SUPPLEMENTO
AL VOLUME VII, SERIE X, DEL
NUOVO CIMENTO
A CURA DELLA SOCIETÀ ITALIANA DI FISICA

1958 — 1° Trimestre — N. 2

RENDICONTI

DEL

V CORSO CHE NELLA VILLA MONASTERO A VARENNA

DAL 14 LUGLIO AL 3 AGOSTO 1957

FU TENUTO A CURA

DELLA SCUOLA INTERNAZIONALE DI FISICA

DELLA SOCIETÀ ITALIANA DI FISICA

INDICE

INTRODUZIONE.

G. POLVANI – Discorso inaugurale . . . . . pag. 159
F. G. FUMI – Prolusione . . . . . » 162

PARTE PRIMA – **Teoria quantistica dei solidi.**

Lezioni

H. BROOKS – Quantum Theory of Cohesion . . . . . pag. 165
D. L. DEXTER – Optical Properties of Solids . . . . . » 245
J. FRIEDEL – Metallic Alloys . . . . . » 287
N. F. MOTT – The Transition from the Metallic to the Non-Metallic State » 312
D. PINES – Electrons and Plasmons . . . . . » 329
J. M. ZIMAN – Transport Properties of Solids . . . . . » 353

Seminari

J. R. SCHRIEFFER – Theory of Superconductivity . . . . . pag. 337

PARTE SECONDA – **Difetti reticolari nei solidi.**

Lezioni

F. C. FRANK – Dislocation Theory . . . . . pag. 386
F. SEITZ – Point Imperfections in Solids . . . . . » 414
A. GUINIER – Effets des imperfections cristallines sur la diffraction des rayons X . . . . . » 444
R. HILSCH und W. MARTIENSSEN: Gitterstörungen an dünnen Schichten » 480
H. PICK – Color Centers in Alkali Halides . . . . . » 498
R. SMOLUCHOWSKI – Radiation Effects in Solids . . . . . » 523
C. ZENER – Anelasticity of Metals . . . . . » 544

Seminari

S. AMELINCKX – Dislocations in Ionic Crystals . . . . . pag. 569
F. C. BROWN – Mobility of Electrons in the Silver Halides . . . . . » 600
Y. HAVEN and J. H. VAN SANTEN – Dielectric Relaxation of Lattice Defects in Crystals . . . . . » 605
T. G. CASTNER W. KÄNZIG, and T. O. WOODRUFF – The Electronic Structure of a *V*-Center . . . . . » 612
A. B. LIDIARD – Correlation Effects in Diffusion in Solids . . . . . » 620
A. SEEGER – On some Physical Consequences of the Non-Linear Theory of Dislocations . . . . . » 632
H. G. VAN BUEREN, J. HORNSTRA and P. PENNING – Dislocations in Ge and Si . . . . . » 646

PARTE TERZA – **Semiconduttori.**

Lezioni

H. Y. FAN – Properties of Semiconductors . . . . . pag. 661
G. BUSCH – Semiconducting Compounds . . . . . » 696
W. KOHN – Shallow Impurity States in Semiconductors . . . . . » 713

Seminari

P. AIGRAIN – Recombination Processes in Semiconductors . . . . . pag. 724
O. MADELUNG – Halbleiter mit kleiner scheinbarer Masse der Ladungsträger . . . . . » 730

La Direzione del Corso ringrazia vivamente i Segretari, dottori FRANCO BASSANI *e* MARIO TOSI, *per la collaborazione prestata nella raccolta degli appunti delle lezioni e nella preparazione dei testi per la stampa.*

INTRODUZIONE

Discorso inaugurale.

DI

G. FOLVANI

Presidente della Società Italiana di Fisica

Contrariamente alle volte precedenti, quest'anno abbiamo voluto dare, all'inaugurazione di questo corso, minore risalto perchè il giorno 27 p. v. sarà qui, in questa sede, ricordata tutta l'attività che l'Ente Villa Monastero ha svolto in questo quinquennio, e ci è sembrato doveroso non tormentare con troppi discorsi i nostri graditi ascoltatori.

Tutto ciò però non esime il Presidente della Società Italiana di Fisica di prendere occasione dalla prima riunione del Corso di quest'anno per dare a tutti voi — docenti, allievi, uditori — qui convenuti per partecipare al Corso, il benvenuto più cordiale da parte della Società Italiana di Fisica e mio personale ed esprimere i ringraziamenti più vivi alle varie persone e ai vari enti che dando, in un modo o in un altro, ma tutti con estrema premura e cortesia, il loro aiuto alla nostra fatica hanno permesso che anche quest'anno il Corso si tenesse. Finanziano il Corso il Ministero della Pubblica Istruzione, il Consiglio Nazionale delle Ricerche, il Comitato Nazionale per le Ricerche Nucleari, la RAI, l'Università di Milano, la Prefettura di Como, le Società Moto Guzzi di Mandello sul Lario, la Società Condor di Milano, le Società Badoni, Caleotto, Carenno, Fiocchi, Metalgraf, Serpentino, Orobia di Lecco; e naturalmente, la Società Italiana di Fisica. A tutti questi enti, ai loro reggitori premurosi e cordiali il ringraziamento più sentito della Scuola e della Società.

Ma accanto agli aiuti finanziari non meno importanti sono quelli scientifici e questi soprattutto ci sono stati porti dai docenti che qui terranno le lezioni: di loro è lungo quest'anno l'elenco dato il particolare carattere che alla organizzazione del Corso ha voluto dare il Direttore prof. FUMI dell'Università di Palermo.

Eccone i nomi: F. G. Fumi, Università di Palermo (Direttore del Corso); P. Aigrain, Ecole Normale Supérieure, Paris; S. Amelinckx, Rijksuniversiteit Gent; H. Brooks, Harvard University, Cambridge Mass.; F. C. Brown, University of Illinois, Urbana; G. Busch, Eidgenössische Technische Hochschule, Zürich; D. L. Dexter, University of Rochester; H. Y. Fan, Purdue University, Lafayette Ind.; F. C. Frank, University of Bristol; J. Friedel, École Nationale Supérieure des Mines, Paris; A. Guinier, Conservatoire National des Arts et Métiers, Paris; Y. Haven, Philips Research Laboratories Eindhoven; R. Hilsch, Universität Göttingen; W. Känzig, General Electric Research Laboratory, Schenectady; W. Kohn, Carnegie Institute of Technology, Pittsburgh Penn.; A. B. Lidiard, Atomic Energy Research Establishment, Harwell; O. Madelung, Siemens-Schuckert-Werke Laboratorium, Erlangen; N. F. Mott, University of Cambridge; H. Pick, Technische Hochschule, Stuttgart; D.Pines, Princeton University; J. R. Schrieffer, University of Illinois, Urbana; A. Seeger, Technische Hochschule, Stuttgart; F. Seitz, University of Illinois, Urbana; R. Smoluchowski, Carnegie Institute of Technology, Pittsburgh Penn.; H. G. van Bueren, Philips Research Laboratories, Eindhoven; C. Zener, Westinghouse Research Laboratories, Pittsburgh Penn.; J. M. Ziman, University of Cambridge.

A tutti quanti ho nominato vada il ringraziamento più vivo della Società Italiana di Fisica e mio personale. Desidero inoltre porgere particolari e affettuose espressioni di gratitudine al prof. Fumi per avere egli con entusiasmo e fattiva abilità organizzato il Corso fino nei più minimi particolari e così bene da rendersi degno di ogni lode, come degna di ogni lode sarà la sua direzione.

E un caloroso e non meno cordiale ringraziamento va anche a voi, allievi e uditori carissimi qui convenuti, insieme con i maestri, da tutte le parti del mondo. E, come gli altri anni, nonostante la lista sia lunga, desidero presentarli reciprocamente.

Allievi: F. Bassani e M. Tosi (Italia), Segretari del Corso; D. Aboav (Inghilterra), A. Ascoli (Italia), M. Asdente (Italia), I. Barducci (Italia), G. Biorci (Italia), G. Boato (Italia), E. Boesman (Belgio), E. Braun (Israele), A. Carlevaro (Italia), R. Colombo (Italia), E. Daniel (Francia), J. des Clozeaux (Francia), A. R. de Vroomen (Olanda), W. Duchateau (Belgio), D. M. Edwards (Inghilterra), R. Feder (Stati Uniti), T. Federighi (Italia), M. Fujimoto (Giappone), W. Gebhardt (Germania), E. Germagnoli (Italia). L. Hedin (Svezia), R. Howard (Stati Uniti), G. V. Kidson (Canadà), L. Koch (Francia), F. A. Levi (Italia), F. Lüty (Germania), A. Manfrino (Italia), G. F. Nardelli (Italia), D. H. Niblett (Inghilterra), N. Nilsson (Svezia), M. Oddo (Italia), A. Paoletti (Italia), L. Passari (Italia), J. K. Roros (Stati

Uniti), H. Samelson (Stati Uniti), W. D. Twose (Inghilterra), C. M. van Baal (Olanda), A. Zmerli (Francia)

Uditori: W. Baltensperger (Svizzera), A. F. Brown (Inghilterra), Ph. Choquard (Svizzera), D. Dautreppe (Francia), P. A. P. Faelens (Belgio), A. J. E. Foreman (Inghilterra), L. Giulotto (Italia), B. B. Goodman (Inghilterra), S. C. Jain (India), C. Lafleur (Belgio), W. Martienssen (Germania), M. Matyas (Cecoslovacchia), A. C. Moore (Inghilterra), P. P. Naor (Israele), W. J. Noble (Canada), G. Picus (Russia), G. W. Pratt (Stati Uniti), P. Pringsheim (Stati Uniti), A. Rgeanov (Russia), A. Rose (Stati Uniti), M. Santangelo (Italia), M. Suffczynski (Polonia), B. Vittoz (Svizzera), J. Wucher (Francia).

* * *

Considerando questa varietà di genti, diverse per cultura, tradizioni, tendenze, idealità e ideologie, qui riunite al solo scopo d'indagare quel maraviglioso libro « dove il senno eterno scrisse i propri concetti », e indagando raggiungere tra loro — e questo avverrà, siatene certi: l'esperienza di questi cinque anni ce lo assicura — quella cordiale colleganza e amicizia che il comune studio e la bellezza e la pace del luogo — anch'esse hanno la loro importanza — fatalmente favoriscono, si è portati a credere che anche questa nostra Scuola, nel suo piccolo, ma in un campo particolarmente sensibile e importante quale è quello dell'alta scienza, possa valere per la sua parte a stabilire o a rafforzare quei legami di reciproca stima e di reciproco rispetto che sono il fondamento di ogni convivenza civile.

Con questa persuasione e con l'augurio che i vostri lavori siano altamente proficui, dichiaro aperto il V Corso della Scuola Internazionale di Fisica della Società Italiana di Fisica.

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Prolusione

DI

F. G. FUMI

Direttore del Corso

This year the International Summer School of the Italian Physical Society is devoted to solid State Physics. The term Solid State Physics in its present connotation was perhaps not adopted till the period of the second world war, and the field it covers is actually very vast, and may appear rather indefinite to an outsider. Indeed one hears that solid state physicists concern themselves with the study of the interactions between electrons, and electrons and phonons in crystalline lattices, problems which have a real taste of theoretical physics, but also with the study of luminescent materials, a field which has almost the sound of technological research. One hears that Solid State Physics includes such widely different branches as superconductivity, semiconductors, the plastic behaviour of metals, radiation damage, ferroelectricity, ferromagnetism. One hears also that Solid State Physics is of very direct use to engineers to develop new materials, and to induce new properties in old materials. This broadness and heterogeneity of the subject prevents, I fear, people outside the field from having a clear idea of what Solid State Physics really is.

In effect, I feel it is only possible to say tautologically that Solid State Physics is the study of the physical properties of solids, particularly of the special properties exhibited by atoms and molecules because of their association and regular periodic arrangement in crystals. The basic unifying feature of Solid State Physics is really represented by the Bravais symmetry of the assembly of atoms in crystalline solids: it is this symmetry which determines, on the one hand, the form of electronic wave functions in solids and the types of atomic vibrations, and, on the other hand, restricts to a few categories the possible types of imperfections in solids. It is the regularity of the crystalline array which makes possible the study of such complicated systems as solids really are, in the same way that the lack of this regularity makes the study of liquids

a more difficult endeavour. Naturally, among the many properties of solids, those tend to be preferred which can be discussed in terms of simple, concrete models, which certainly simplify the real problem a great deal, but include its essential elements. This modellistic character of Solid State Physics is perhaps its most noteworthy feature.

My aim in planning this Summer Course has been to give a broad outline of the different modern aspects of Solid State Physics, to illustrate how they all actually link together. I focused my attention, on the one hand, on the more theoretical aspects, the quantum theory of solids, and, on the other hand, on the imperfection-determined properties of solids, which are receiving such world-wide attention. The field of semiconductors, the third main line of the course, can in some ways be envisaged as a cross-product of the other two, but it really has an individuality of its own. The major field of Solid State Physics which was consciously left out of the program is the field of magnetic properties of solids which received detailed attention in the very successful course that Prof. GIULOTTO organized at Varenna last summer.

On the quantum theory of solids, we shall have six series of lectures various and one seminar. Prof. BROOKS will deal in his seven lectures with the basic concepts of energy bands in solids, and with the approximational techniques to deal with the problem of cohesion in solids. Prof. DEXTER will treat the optical properties of solids, dealing in particular with recent views on the exciton. Prof. FRIEDEL will discuss metallic alloys in the light of quantum theory. Prof. MOTT will examine the problem of the transition from the insulating to the conducting state in a solid. Prof. PINES will devote five lectures to a presentation of the method of collective coordinates that he and Prof. BOHM developed a few years ago, illustrating applications of it to the study of electron-electron and electron-phonon interactions in solids. The electron-phonon interactions will then be discussed in more detail by Dr. SCHRIEFFER in his seminar on the recent Bardeen-Cooper-Schrieffer theory of superconductivity. Finally Prof. ZIMAN will discuss in three lectures the effect of lattice imperfections on the thermal and electrical conductivity of solids.

On lattice imperfections specifically, we will have seven series of lectures various and eight seminars. Prof. FRANK will illustrate the basic concepts of dislocation theory, while Prof. SEITZ will treat point imperfections and their role in the physics of solids in general. Prof. GUINIER will illustrate the type of information that X-ray techniques can give on the presence of lattice imperfections. Prof. HILSCH will speak on the recent Göttingen work on thin films, and Prof. PICK will discuss classical and recent work on the most important types of colour centres. Prof. SMOLUCHOWSKI will present the recent ideas on the yet somewhat puzzling, and certainly involved, pheno-

menon of radiation damage. Finally Prof. ZENER will illustrate various applications of the anelastic techniques that he has brought into Solid State Physics with such success. The titles of the individual seminars on imperfections are, I feel, self-explanatory.

Concerning semiconductors, I should first say how sorry we all are that Dr. SHOCKLEY, Nobel Prize winner, who was scheduled to give a group of five lectures on transistor physics, could not be with us owing to urgent needs of his industrial laboratory in California. However, we are fortunate to have here Prof. FAN who will give a group of seven introductory lectures on the bulk properties of valence semiconductors. Prof. BUSCH will then discuss the recent work on the new types of semiconducting compounds, such as InSb, and Prof. KOHN will treat the important problem of impurity levels in semiconductors. Dr. AIGRAIN and Dr. MADELUNG will also give two seminars.

The program as it stands probably looks rather formidable, but I have strong hopes that the relaxing atmosphere of Lake Como, and the friendliness and informality that I am sure will permeate our meetings will help us to survive it.

* * *

Per concludere, sono certo di esprimere il pensiero di tutti i partecipanti al corso porgendo il più sentito ringraziamento all'Ente Villa Monastero e al suo Presidente avv. BOSISIO per l'ospitalità concessaci, e un particolare ringraziamento alla Società Italiana di Fisica, specie nella persona del suo instancabile Presidente prof. POLVANI, per aver deciso di dedicare questo 5° Corso della Scuola Internazionale di Fisica all'argomento che appassiona tutti noi, e con noi varie migliaia di fisici in tutto il mondo: la Fisica dei solidi.

SOCIETÀ ITALIANA DI FISICA

SCUOLA INTERNAZIONALE DI FISICA

5° CORSO ESTIVO - VARENNA SUL LAGO DI COMO - VILLA MONASTERO - 14 Luglio - 3 Agosto 1957

1. Mrs. Feder
2. M. Rumignani
3. R. Howard
4. A. Rzhanov
5. E. Germagnoli
6. Mrs. Howard
7. G. Pikus
8. W. Kohn
9. E. Boesman
10. C. Zener
11. H. Samelson
12. D. H. Niblett
13. R. Smoluchowski
15. W. D. Twose
16. H. Brooks
17. I. Barducci
18. A. Manfrino
19. F. Seitz
20. W. J. Noble
21. Ph. Choquard
22. N. F. Mott.
23. M. Suffczynski
24. L. Hedin
25. Mrs. Pratt
26. W. Baltensperger
27. J. Wucher
29. F. C. Frank
30. G. V. Kidson
31. A. R. de Vroomen
32. C. M. van Baal
33. S. C. Jain
34. H. Y. Fan
35. P. P. Naor
36. G. Biorci
37. L. Ballerini
38. E. Braun
39. G. W. Pratt
40. D. Pines
41. M. Oddo
43. L. Koch
44. A. F. Brown
45. E. Daniel
46. M. Matyas
47. O. Madelung
48. F. A. Levi
49. M. Fujimoto
50. M. Asdente
51. R. Colombo
52. G. Boato
53. L. Passari
54. F. Lüty
55. A. Zmerli
57. D. M. Edwards
58. P. A. P. Faelens
59. J. des Cloizeaux
60. G. F. Nardelli
61. W. Duchateau
62. Mrs. Tosi
63. M. Tosi
64. Mrs. Polvani
65. A. J. E. Foreman
66. Y. Haven
67. D. Aboav
68. A. Paoletti
69. P. Giacometti
71. D. Dautreppe
72. Mrs. Dautreppe
73. Mrs. Frank
74. Mrs. Fumi
75. F. G. Fumi
76. A. Ascoli
77. G. Polvani
78. A. Carlevaro
79. A. Federighi
80. Mrs. Zener
81. N. Nilsson
82. Mrs. Fan

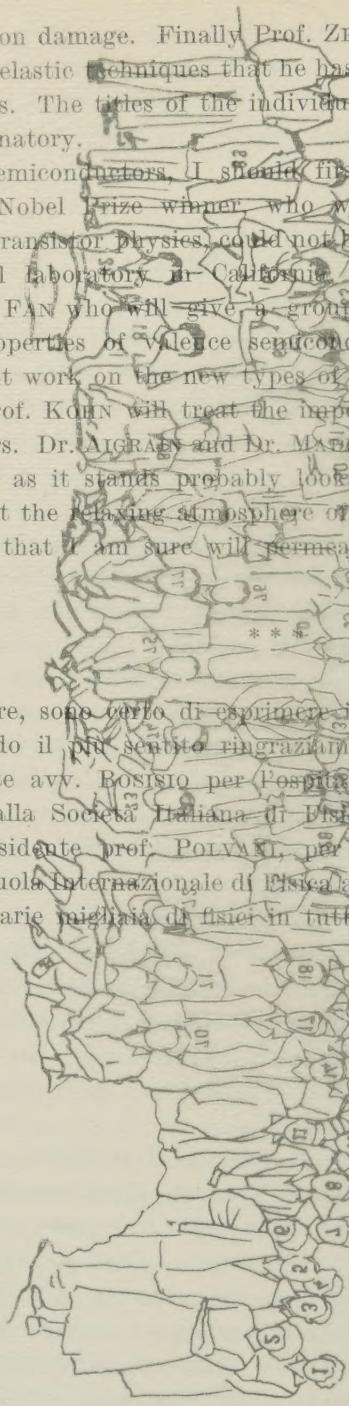
menon of radiation damage. Finally Prof. ZENER will illustrate various applications of the anelastic techniques that he has brought into Solid State Physics with such success. The titles of the individual seminars on imperfections are, I feel, self-explanatory.

Concerning semiconductors, I should first say how sorry we all are that Dr. SHOCKLEY, Nobel Prize winner, who was scheduled to give a group of five lectures on transistor physics, could not be with us owing to urgent needs of his industrial laboratory in California. However, we are fortunate to have here Prof. FAN who will give a group of seven introductory lectures on the bulk properties of valence semiconductors. Prof. BUSCH will then discuss the recent work on the new types of semiconducting compounds, such as InSb, and Prof. KOHN will treat the important problem of impurity levels in semiconductors. Dr. ALLGAIER and Dr. MANLING will also give two seminars.

The program as it stands probably looks rather formidable, but I have strong hopes that the relaxing atmosphere of Lake Como, and the friendliness and informality that I am sure will permeate our meetings will help us to survive it.

* * *

Per concludere, sono certo di esprimere il pensiero di tutti i partecipanti al corso porgendo il più sentito ringraziamento all'Ente Villa Monastero e al suo Presidente avv. BOSISIO per l'ospitalità concessaci, e un particolare ringraziamento alla Società Italiana di Fisica, specie nella persona del suo instancabile Presidente prof. POLVANI, per aver deciso di dedicare questo 5° Corso della Scuola Internazionale di Fisica all'argomento che appassiona tutti noi, e con noi varie migliaia di fisici in tutto il mondo: la Fisica dei solidi.



SOCIETÀ ITALIANA DI FISICA

SCUOLA INTERNAZIONALE DI FISICA

5° CORSO ESTIVO - VARENNA SUL LAGO DI COMO - VILLA MONASTERO - 14 Luglio - 3 Agosto 1957

PARTE PRIMA

Teoria quantistica dei solidi.

Quantum Theory of Cohesion.

H. Brooks

Harvard University, Cambridge, Massachusetts

I.

Qualitative considerations.

1. – Introduction.

The cohesive energy of a crystal is defined as the difference in total energy between the crystal and an equal number of separated neutral atoms. Experimentally it may be determined as the heat of sublimation.

In these lectures we shall be concerned almost exclusively with the cohesion of simple metals, and shall not discuss ionic or covalent solids in any detail. The subject of cohesive energy really embraces three closely related topics:

1) The value of the cohesive energy at the equilibrium lattice constant.

2) The variation of cohesive energy with lattice constant, i.e. experimentally the pressure-volume relations at absolute zero.

3) The prediction of the equilibrium lattice constants.

A more difficult problem is the prediction of the relative stability of allotropic modifications. The differences in energy between such modifications usually are less than 3 kcal per mol, and, except for a few qualitative ideas, are beyond the scope of present quantitative theory. Since these lectures will be primarily concerned with the quantitative theory of cohesion, we will not touch on the subject of phase changes. In the present state of the theory we can discuss quantitatively only those factors in the cohesive energy which depend upon the atomic volume and are relatively independent of the arrangement of atoms in the crystal structure.

We may divide the cohesive energy of a metallic element into two parts, as follows:

1) A part arising from the fact that the free atom wave functions must be modified so as to join smoothly from cell to cell in the crystal.

2) A part which arises from the interactions of the overlapping electrons among each other.

The first energy has been termed by WIGNER and SEITZ [1] the boundary correction. To a certain approximation it may be regarded as a one-electron problem. It is the part of the problem which is concerned with the formation of energy bands in crystals, and will be the major concern of these lectures.

The second part of the energy is usually referred to as the interaction energy. It is the part of the problem of cohesion which is specifically a many electron problem. In these lectures we shall have little to say in detail about the interaction energy, since this will form a major topic of Professor PINES's lectures.

2. – Units and nomenclature.

In these lectures we shall use the Bohr system of units in which the unit of energy is the rydberg (1 Ry = 13.60 eV), and the unit of length is the radius ($0.5292 \cdot 10^{-8}$ cm) of the first Bohr orbit in a hydrogen atom of infinite nuclear mass. In this system of units we should note particularly that the Schrödinger equation of a hydrogen atom takes the form

$$\nabla^2\psi + \left[E + \frac{2}{r}\right]\psi = 0\,. \tag{1.1}$$

The potential of a point charge is $2/r$, and the energy associated with a plane wave of wave vector $\boldsymbol{k}$, i.e. of the form $\psi = \exp[i\boldsymbol{k}\cdot\boldsymbol{r}]$, is $E = k^2$. The cohesive energy may be expressed in units of rydbergs per atom. Often cohesive energies are expressed in kcal per mole, the conversion being 1 Ry per atom = = 312 kcal/mol.

3. – Alkali metals.

These are the simplest of all metallic solids principally because they may be described in terms of one electron per atom outside of well-defined ion cores which do not at all participate in binding and are in fact unmodified in the solid. This rigorous separation of ion cores and valence electrons is exemplified in the large difference in binding energy between the most tightly bound valence electron in the metal and the most weakly bound electron in the core. These two energies and their ratio are shown in Table I, all in rydbergs.

TABLE I.

Element	Valence Electron	Core Electron	Core/Valence
Li	.688	5.56	8.1
Na	.606	3.60	5.9
K	.494	2.34	4.7
Rb	.464	2.04	4.4
Cs	.429	1.73	4.0

For all these metals the loosest binding to the core is between 4 and 8 times the tightest binding of a valence electron in the metal.

It is also of interest to compare the cohesive energy of the metal with the ionization energy of the free atom. This is shown in Table II, which indicates that the cohesive energy is between $\frac{1}{5}$ and $\frac{1}{4}$ the ionization energy with surprising constancy for all the elements except Li.

TABLE II.

		Li	Na	K	Rb	Cs
Ratio	$\frac{\text{Cohesive Energy}}{\text{Ionization Energy}}$	.295	.227	.218	.214	.211

4. – Multivalent metals with valence electrons outside rare gas configuration.

In these elements the separation between valence electrons and ion cores can be made as in the alkali metals. Indeed, because of the nuclear charge the rare gas ion cores are even more compact than in the alkali metals, and hence more completely isolated from binding in the metal. It is interesting to observe that when cohesion is expressed in terms of binding per valence electron, binding energies are of the same order of magnitude for most of the elements in this class. As in the case of the alkali metals the elements from the first row of the periodic table show exceptionally high binding. In Table III we show cohesive energies per electron for elements of valence 1, 2, 3 and 4.

TABLE III.

Li	.117	Be	.123	B	.103	C	.137
Na	.084	Mg	.058	Al	.079	Si	.070
K	.070	Ca	.074	Sc)	.099	Ti)	.090
Rb	.066	Sr	.063	Y)	.110	Zr)	.100
Cs	.060	Ba	.067	La)	.094	Ce)	.068

The elements followed by) are of doubtful significance in the table because with these elements the valence electrons in the free atom are beginning to fill a *d* or *f* shell. In the solid it is probable that there is a corresponding *d* or *f* band which is low lying and which partially overlaps the normal valence band, thus increasing the cohesive energy because the valence electrons have less average kinetic energy. It is interesting to notice how little the general trends are affected by the fact that some of the elements tabulated are insulators or semiconductors rather than metals. The modifications of the band structure which are responsible for profound changes in their electrical and optical properties have little or no influence on the magnitude of the cohesive energy, although they do have a bearing on what crystal structure is stable.

5. – Monovalent noble metals: Cu, Ag, Au.

These metals are very similar to the alkalis, except that the single valence electron now is outside a completed *d* subshell rather than a rare gas configuration. The resulting ion core is much more loosely bound in comparison with the valence electron than is the case for the alkali metals. In Cu, for example, the lowest state of a valence electron in the metal is bound about 1 Ry, while the loosest core electron is bound by only 1.5 Ry. Thus in the solid there is really very little distinction between ion core and valence electrons. As a corollary of this the ion cores are very large. In the elements discussed in Sect. 3 and 4 the equilibrium atomic volume is essentially determined by the pressure of the Fermi gas formed by the valence electrons, and the cores are very far apart. In Cu, Ag and Au, by contrast, the lattice spacing is governed by the overlap of the ion cores. The interaction between the valence electrons and the ion cores is much stronger than in the alkali metals, because the nuclear charge is less well shielded from the valence electrons by the ion core. In consequence, the equilibrium atomic volume resulting from the compromise between valence-electron ion-core interaction (attractive) and the pressure of the Fermi gas (repulsive) would be much smaller than the observed constant were it not for the overlap of the cores.

This very different character of the ion-core valence-electron relationship in the monovalent noble metals is reflected in very different magnitudes of cohesion, as indicated in Table IV.

TABLE IV.

Metal	Cohesion	Ratio Cohesion/Ionization Energy
Cu	.260 Ry	.46
Ag	.221 »	.39
Au	.264 »	.43

The cohesion per valence electron is 3 to 4 times what it is in other metals, and even the ratio to the ionization energy is about twice what it is in other cases. Detailed calculations, which will be mentioned in the last lecture, show that the interaction between valence electron and ion core falls far short of accounting for this large binding. Thus we must assume that the closed *d* shells contribute significantly to the binding.

6. – Multivalent metals with valence electrons outside completed *d* subshells.

The most striking feature here is the sharp decrease in binding as we move to the right in the periodic table from Cu, Ag or Au. This is most striking when we tabulate cohesion per valence electron as shown in Table V.

TABLE V.

Cu	.260	Zn	.050	Ga	.071	Ge	.063
Ag	.221	Cd	.043	In	.062	Sn	.058
Au	.264	Hg	.024	Tl	.046	Pb	.037

It is immediately evident that the metals in columns 2, 3 and 4 show near normal binding per electron in comparison with the noble metals. This is essentially because the ion core again becomes isolated as we increase the nuclear charge. Thus as we go from Cu to Zn the extra nuclear charge contracts the *d* shell drastically. This both decreases the penetration of the valence electron into the core, thus decreasing its interaction with the nucleus, and decreases the polarizability of the *d*-shell, thus removing the possibility of a strong interaction between *d*-shells. Quantitative calculations indicate that for metals like Zn, in contrast to Cu, the interaction between valence electron and core is sufficient to account for the observed cohesion.

7. – Transition metals.

We have seen that in the noble metals, even the closed *d*-shells may contribute substantially to cohesion. The contribution is even greater if the *d*-shells are not complete, as in the transition groups. The usual explanation of this goes roughly as follows. The *d*-shells overlap in the solid and are therefore broadened into bands; for Cu, Ag and Au these must be filled bands, and in such a case it can be shown that to a fairly good approximation the average energy of the electrons in the band is the same as the original atomic level. In this approximation there should thus be no bonding due to *d*-electrons.

It is thus probable that for the noble metals bonding is essentially a many-electron effect, arising from correlations in the motion of the electrons in adjacent *d*-shells. When the *d*-shell is incomplete, however, an additional contribution to cohesion results from the fact that the electrons fill up only the lower states of the band. According to this argument the maximum cohesion in the transition elements should be reached when the *d*-shell is near to half filled, which is observed to be the case. The maximum total cohesive energies in each of the transition groups are shown in Table VI, together with

TABLE VI.

Element	Total Cohesion	Cohesion per electron
V	.383 (5)	.077
Nb	.591 (5)	.118
W	.646 (6)	.108

the binding energy per electron computed on the basis of all the electrons outside the nearest rare gas configuration. The number in ( ) is the number of valence electrons used in the calculation.

It is interesting to observe that the binding energy per electron is of the same general order of magnitude as for the other multivalent elements. That it is somewhat on the high side is probably the result of the fact that we are dealing with *d*-electrons which have a high degree of degeneracy, and therefore do not have as high a kinetic energy on the average as normal valence electrons.

8. – Other elements.

In the case of elements to the right of the 4th column of the periodic table we find that the cohesion per electron *lacking* from a complete rare gas shell is fairly constant. This point is illustrated by Table VII. In this case the result is not so surprising, for we know that all these elements tend to form structures in which the co-ordination number is equal to the number of holes

TABLE VII. – *Energy per hole.*

F	.057 Ry	O	.094	N	.059
Cl	.112	S	.086	P	.080
Br	.086	Sc	.077	As	.065
I	.088	Te	.076	Sb	.065
				Bi	.053

per atom. Thus the halogens form diatomic molecular solids in which each atom is tightly bonded to only one neighbor. S, Se and Te form complex structures in which the atoms are arranged in chains with each atom having two neighbors closer than the rest. As, Sb and Bi are semimetals and tend to form structures of layer-like character in which the atoms are arranged in buckled planes with co-ordination 3. All of these crystals may be described qualitatively in terms of chemical bonds between definite pairs of atoms, one bond for each hole per atom, and the cohesion is roughly describable in terms of energy per bond.

There exists a special class of solids, the noble gases (or rare gases) in which there are no valence electrons. These show very weak cohesion compared with the other elements discussed above. Such cohesion as there is results entirely from the very weak van der Waals or polarization interaction between the closed shells. This does not result from overlap of the wave functions, but from correlations in the motions of the electrons in adjacent shells through their Coulomb interactions. The cohesion of the noble gases is shown in Table VIII.

Table VIII.

He	.00008 Ry
Ne	.0016 »
A	.0059 »
Kr	.0081 »
Xe	.0114 »
Rn	.0147 »

REFERENCES

[1] E. P. Wigner and F. Seitz: *Qualitative Analysis of the Cohesion of Metals*, pp. 97-126 in *Solid State Physics*, vol. I, ed. by F. Seitz and D. Turnbull: (New York, 1955). Cfr. also E. P. Wigner: *Proc.* 1953 *Intern. Conf. Theor. Phys.* (Tokyo, 1954), p. 649.

[2] An earlier qualitative discussion of cohesion is given in F. Seitz: *Modern Theory of Solids* (New York, 1940), chap. I, pp. 1-75. Cfr. also C. Kittel: *Introduction to Solid State Physics*, 2nd ed. (New York, 1956), chap. 3, pp. 63-85.

[3] For a general discussion of cohesion of metals, see *Theory of Alloy Phases, American Society for Metals* (Cleveland, 1956), especially the chapters by J. C. Slater, pp. 1-13, and H. Brooks, pp. 199-220.

[4] The numerical values of cohesive energy are taken from Table II, p. 108 of ref. [1].

II.

A solid as a many-electron system.

1. – The density matrix.

Suppose that we have a system containing N electrons which may be described by the many-electron wave function

$$\Psi(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N) ,$$

where $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ..., $\boldsymbol{x}_N$ symbolize position vectors for the N electrons which also include the spin variables. We may define the first and second order density matrices as follows, using a notation due to LÖWDIN [1]:

$$\Gamma(\boldsymbol{x}_1) = N \int |\Psi|^2 \, \mathrm{d}\boldsymbol{x}_2 \ldots \mathrm{d}\boldsymbol{x}_N , \tag{2.1a}$$

$$\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{N(N-1)}{2} \int |\Psi|^2 \, \mathrm{d}\boldsymbol{x}_3 \ldots \mathrm{d}\boldsymbol{x}_N , \tag{2.1b}$$

where the integrations are taken over all the electronic co-ordinates except those which occur explicitly in the argument of the Γ's, and summation over spin variables is understood. It is also necessary to define the joint density matrix

$$\Gamma(\boldsymbol{x}_1' | \boldsymbol{x}_1) = N \int \Psi^*(\boldsymbol{x}_1', \boldsymbol{x}_2, \ldots \boldsymbol{x}_N) \Psi(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_N) \, \mathrm{d}\boldsymbol{x}_2 \ldots \mathrm{d}\boldsymbol{x}_N , \tag{2.2}$$

in which the wave functions in the integrand differ in the argument $\boldsymbol{x}_1$, i.e. in the position of electron 1.

In order to appreciate the physical meaning of these definitions we consider a simple illustration. Suppose that the wave function may be written as a simple Slater determinant, or antisymmetrized sum of products, a typical product being

$$\psi_{K_1}(\boldsymbol{x}_1) \psi_{K_2}(\boldsymbol{x}_2) \ldots \psi_{K_N}(\boldsymbol{x}_N) \tag{2.3}$$

then we have

$$\Gamma(\boldsymbol{x}_1' | \boldsymbol{x}_1) = \sum_{K=1}^{N} \psi_K^*(\boldsymbol{x}_1') \psi_K(\boldsymbol{x}_1) \tag{2.4}$$

and

$$\Gamma(\boldsymbol{x}_1) = \sum_{K=1}^{N} |\psi_K(\boldsymbol{x}_1)|^2 . \tag{2.5}$$

where the summation is taken over all the states in the determinant. For the second order density matrix we also have

$$\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2) = \tfrac{1}{2}\Gamma(\boldsymbol{x}_1)\Gamma(\boldsymbol{x}_2) - \tfrac{1}{2}|\Gamma(\boldsymbol{x}_1|\boldsymbol{x}_2)|^2 . \tag{2.6}$$

The quantity $\Gamma(\boldsymbol{x}_1)$ represents an ordinary charge density (always including spin) while $\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is proportional to the probability of finding one electron at $\boldsymbol{x}_1$ and another at $\boldsymbol{x}_2$ simultaneously.

Using the above general expressions for the density matrices the expectation value for the energy of a solid may be written in the form:

$$E = E_1 + E_2 + E_3 , \tag{2.7}$$

where

$$E_1 = \int p_1^2 \Gamma(\boldsymbol{x}_1'|\boldsymbol{x}_1)\delta(\boldsymbol{x}_1 - \boldsymbol{x}_1')\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}\boldsymbol{x}_1' , \tag{2.8a}$$

$$E_2 = \sum_{gh}{}' \frac{Z_g Z_h}{r_{gh}} - 2\sum_g Z_g \int \frac{\Gamma(\boldsymbol{x}_1)}{r_{1g}}\,\mathrm{d}\boldsymbol{x}_1 + \int \frac{\Gamma(\boldsymbol{x}_1)\Gamma(\boldsymbol{x}_2)}{r_{12}}\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}\boldsymbol{x}_2 , \tag{2.8b}$$

$$E_3 = 2\int \frac{\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2) - \tfrac{1}{2}\Gamma(\boldsymbol{x}_1)\Gamma(\boldsymbol{x}_2)}{r_{12}}\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}\boldsymbol{x}_2 . \tag{2.8c}$$

In these expressions p_1^2 is the kinetic energy operator, operating on the variable $\boldsymbol{x}_1$. The term E_1 represents the expectation value of the kinetic energy of the system. In E_2 the $'$ in the first summation indicates omission of the term for $g = h$ in the sum over lattice points. r_{gh} is the distance between lattice points labelled g and h, r_{1g} is the distance between lattice point g and electron 1, r_{12} is the distance between electrons 1 and 2.

The 3 parts of the energy in (2.8) have been separated in such a way as to permit pairing off terms to eliminate divergences in the sums. The convergence difficulties are all concentrated in the E_2 term. E_2 is sometimes referred to as the classical electrostatic energy or Coulomb energy, in that it is calculated by purely classical or electrostatic arguments, but with charge densities computed quantum mechanically. However, it is a classical calculation only in the sense that it involves only the first order charge densities. E_3 is the quantum mechanical part of the electrostatic energy, in that it involves the joint probability density. It gives rise to what are usually termed the exchange and correlation energies.

2. – Simplification of E_1.

If the wave function can be written as a Slater determinant, E_1 becomes just the sum of the kinetic energies of the individual electrons, including the core electrons. In this case it can be written in the form

$$E_1 = \sum_K \int \psi_K^*(\boldsymbol{x}_1) p_1^2 \psi_K(\boldsymbol{x}_1)\, \mathrm{d}\boldsymbol{x}_1 = -\sum_K \int \psi_K^*(\boldsymbol{x}_1 \nabla^2 \psi_K(\boldsymbol{x}_1)\, \mathrm{d}\boldsymbol{x}_1 \,. \tag{2.9}$$

3. – Simplification of E_2 and E_3.

The individual terms in (2.8*b*) are divergent in the sense that they give contributions to the energy which increase faster than proportionally to the number of atoms in the crystal. In fact each term increases as $N^{8/3}$, where N is the number of atoms in the crystal. Despite this, if the crystal is assumed to be electrically neutral the divergent parts of each term cancel against each other, and we find that E_2 increases proportionally to N.

We assume that we can write

$$\Gamma(\boldsymbol{x}_1) = \sum_g \gamma_g(\boldsymbol{x}_1) \,, \tag{2.10}$$

where $\gamma_g(\boldsymbol{x}_1)$ is a localized charge density associated with the cell centered at g. There are an infinite number of ways of performing the decomposition of the charge density in (2.10). For convenience we will assume that $\gamma_g(\boldsymbol{x}_1) = \Gamma(\boldsymbol{x}_1)$ within the cell g and vanishes elsewhere.

If the point $\boldsymbol{x}_2$ lies outside the cell h we may write

$$\int \frac{\gamma_h(\boldsymbol{x}_1)}{r_{12}}\, \mathrm{d}\boldsymbol{x}_1 = \frac{n_h}{r_{2h}} - \omega_h(\boldsymbol{x}_2) \,, \tag{2.11}$$

where

$$n_h = \int \gamma_h(\boldsymbol{x}_1)\, \mathrm{d}\boldsymbol{x}_1 \,. \tag{2.12}$$

Here n_h is seen to be just the electronic charge associated with the h-th cell. The only property of $\omega_h(\boldsymbol{x}_2)$ which we need to use is that it falls off faster than $1/r_{2h}$. It represents the effect on the potential of the dipole and higher moments of the charge distribution $\gamma_h(\boldsymbol{x}_1)$. We may also write

$$\int \frac{2\gamma_g(\boldsymbol{x}_1)}{r_{1g}}\, \mathrm{d}\boldsymbol{x}_1 = \frac{2n_h}{r_{gh}} - 2\omega_h(g) \,, \tag{2.13}$$

where the right hand side is the potential produced at the nucleus g by the electronic charge distribution associated with cell h.

Using (2.10), (2.11), and (2.12) in the expression for E_2 we obtain

$$E_2 = \sum_{gh}{}' \frac{(Z_g - n_g)(Z_h - n_h)}{r_{gh}} + \\ + \sum_{gh}{}' \left\{(2Z_g - n_g)\omega_h(g) - \int \gamma_g(\boldsymbol{x}_1)\omega_h(\boldsymbol{x}_1)\,\mathrm{d}\boldsymbol{x}_1\right\} + \\ + \sum_{g} \left\{-2Z_g \int \frac{\gamma_g(\boldsymbol{x}_1)}{r_{1g}}\,\mathrm{d}\boldsymbol{x}_1 + \int \gamma_g(\boldsymbol{x}_1) \int \frac{\gamma_g(\boldsymbol{x}_2)}{r_{12}}\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}\boldsymbol{x}_2\right\}. \tag{2.14}$$

In this expression the first term is the Madelung energy [2]. Its convergence can be demonstrated from the fact that

$$\sum_g n_g = \sum_g Z_g\,, \tag{2.15}$$

i.e. the total electronic charge is equal to the total nuclear charge. For elemental metals with which we shall be concerned all the n's and Z's are the same, so that the Madelung energy vanishes.

The second term in (2.14) is a consequence of the lack of spherical symmetry of the charge distribution associated with each cell. It may be written in the form

$$\sum_{gh}{}' \left\{(Z_g - n_g)\,\omega_h(g) + \int [Z_g\,\delta(\boldsymbol{x}_1) - \gamma_g(\boldsymbol{x}_1)]\,\omega_h(\boldsymbol{x}_1)\,\mathrm{d}\boldsymbol{x}_1\right\}. \tag{2.16}$$

The first term may be thought of as the interaction of the non-spherically symmetric part of the charge distribution in cell h with the potential produced by the net charge in cell g, while the second term may be thought of as the interaction of the total charge distribution in cell g (including the nuclear charge) with the potential produced by the non-spherically symmetric part of the charge distribution in cell h. Again the neutrality of the crystal guarantees the convergence of these terms. For simple metals this contribution to the energy is probably very small. Seitz [3] has estimated the total contribution of deviations from spherical symmetry to the energy of a BCC array of positive ions embedded in a uniform distribution of negative charge, and found that it amounts to only 0.001 Ry in the case of Na, contributing a slight increase in cohesion. Seitz's estimate includes not only the second term of (2.14) but also contributions from the non-spherical parts of the γ's in the 3rd term.

The third term of (2.4) arises from the interactions of the charge distributions within the same cell. The first term in the bracket is the self energy

of the electron distribution within one cell. This second term is what is usually referred to in the literature as the Coulomb energy of the electrons. This terminology is slightly misleading because it implies that this represents the whole electrostatic energy of the electrons. We have seen that all the terms in E_2 contain electronic interactions, but that these other terms are almost precisely cancelled by electrostatic interactions between nuclei, or between electrons and nuclei.

If the wave functions can be written as Slater determinants, then we may make a separation of $\gamma_g(\boldsymbol{x}_1)$ into a core part $\gamma_g^{(c)}(\boldsymbol{x}_1)$ and a valence part $\gamma_g^{(v)}(\boldsymbol{x}_1)$. We may further assume, as is valid for many metals, that the core part is unaltered in the solid.

It is convenient to write (2.8c) in the form:

$$E_3 = \int \frac{\Gamma(\boldsymbol{x}_1)\left\{\begin{matrix}2\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ \Gamma(\boldsymbol{x}_1)\end{matrix} - \Gamma(\boldsymbol{x}_2)\right\}}{r_{12}} \mathrm{d}\boldsymbol{x}_1 \mathrm{d}\boldsymbol{x}_2 . \tag{2.17}$$

This may be considered as representing the interaction of a charge distribution $\Gamma(\boldsymbol{x}_1)$ with the charge distribution represented by the quantity in brackets. From the definitions of $\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and of $\Gamma(\boldsymbol{x})$ it is readily shown that the total charge density corresponding to the bracket is given by

$$\int \left\{\begin{matrix}2\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ \Gamma(\boldsymbol{x}_1)\end{matrix} - \Gamma(\boldsymbol{x}_2)\right\} \mathrm{d}\boldsymbol{x}_2 = -1 , \tag{2.18}$$

which is just the negative of 1 electron. Thus the charge distribution in (2.18) represents a deficiency of electronic charge, whose distribution depends upon the position of the first electron. For large distances between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ it is plausible that the probability of finding one electron is independent of the position of the other electron, which implies that

$$\frac{2\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2)}{\Gamma(\boldsymbol{x}_1)} = \Gamma(\boldsymbol{x}_2) , \tag{2.19}$$

so that the integrand of (2.18) vanishes when electron 2 is distant from electron 1. Thus the bracketed expression in (2.18) may be thought of as representing a « hole » whose total strength is one electronic charge and whose maximum depth tends to follows electron 1 around as it moves. The interaction of electron 1 with its accompanying hole gives a negative contribution to the total energy, and since $\Gamma(\boldsymbol{x}_1)$ is the total charge distribution due to N electrons we see that E_3 is proportional to N. For determinantal wave func-

tions our « hole » just reduces to the Fermi or exchange hole which has been frequently discussed in the literature [4] but the above discussion shows that its significance is more general. It is generally found that the size of the hole is of the order of magnitude of the spacing between electrons in the electron gas formed by the valence electrons. As a result, in a monovalent metal the energy E_3 nearly cancels with the Coulomb self energy of the last term in E_2. This is illustrated in Table IX, which shows for the alkali metals the contribution to the total energy per atom of the sum of the last term of E_2 and the E_3 contribution. In computing E_3 we used the theory of correlation energy which Professor PINES will discuss in his lectures. Table IX actually shows

TABLE IX. – *Interaction Energy.*

Li	.0127 Ry	.0125
Na	.0030 »	.0069
K	.0035 »	.0039
Rb	.0051 »	.0032
Cs	.0062 »	.0029

two results, the second column representing the result found using the latest refinements in Pines's theory.

In the case of polyvalent metals the correlation hole tends to be somewhat smaller than the unit cell, so that the Coulomb contribution to the energy predominates. This situation is illustrated in Table X, which shows the total

TABLE X. – *Interacting Energy.*

Na	.0069 Ry
Mg	.2871 »
Al	.6702 »
Si	.9498 »

of the interaction energies for Na, Mg, Al and Si computed on the assumption that the electrons may be treated as free. The energies are expressed per valence electron rather than per atom in order to provide a better comparison with Na. The new Pines formula is used. For reference we give both the old [5] and new formulas in the equation below:

$$\text{(2.20)} \qquad \text{Old:} \quad E_{\substack{\text{corr.}\\ \text{exch.}}} = -\frac{.916}{r_s} - .114 + .0313 \ln r_s + .0005 r_s ,$$

$$\text{(2.21)} \qquad \text{New:} \quad E_{\substack{\text{corr.}\\ \text{exch.}}} = -\frac{.916}{r_s} - .140 + .0626 \ln r_s - .0028 r_s ,$$

where it must be remembered that r_s represents not the Wigner-Seitz sphere radius but the radius of the sphere whose volume is equal to the volume per electron, i.e. the reciprocal of the electronic density.

If the wave function is a Slater determinant, then it is readily verified that

$$2\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2) - \Gamma(\boldsymbol{x}_1)\Gamma(\boldsymbol{x}_2) = -|\Gamma(\boldsymbol{x}_1|\boldsymbol{x}_2)|^2 , \tag{2.22}$$

where $\Gamma(\boldsymbol{x}_1|\boldsymbol{x}_2)$ is the mixed probability density defined by (2.4) and is not to be confused with $\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2)$ itself. If we introduce spin variables explicitly we can separate $\Gamma(\boldsymbol{x}_1|\boldsymbol{x}_2)$ into an up spin and a down spin part. If we use this separation in the expression for E_3 and remember that the co-ordinate integrations must include a sum over spin variables, it is immediately apparent that the cross terms between up and down spin densities go out. We may thus write

$$E_3 = -\int \frac{|\Gamma_\uparrow(\boldsymbol{x}_1|\boldsymbol{x}_2)|^2 + |\Gamma_\downarrow(\boldsymbol{x}_1|\boldsymbol{x}_2)|^2}{r_{12}} \mathrm{d}\boldsymbol{x}_1 \mathrm{d}\boldsymbol{x}_2 . \tag{2.23}$$

This may be interpreted by saying that the correlation hole charge density exists only among the electrons having spin parallel to the given electron. There is no correlation between the positions of electrons of opposite spin. It is to be noted that the total strength of the hole is still -1 electron. Thus with determinantal wave functions we have correlation between electrons of parallel spin but none between electrons of antiparallel spin. This arises essentially because for antisymmetric wave functions we must have

$$\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0 , \tag{2.24}$$

i.e. the probability of two electrons with parallel spin occupying the same point in space is zero, a fact which is a direct expression of the Pauli exclusion principle. In a more exact treatment there is correlation between electrons of both spin directions, but the total charge connected with the correlation hole remains the same. That the total energy E_3 is more negative when correlation is taken into account must be a consequence of the fact that the distribution of charge in the hole is less extended in this case so that the interaction of the electron with its hole is greater. This is a natural consequence of the fact that the exchange hole is only half as deep at its center as is the combined exchange and correlation hole, as can be seen immediately from the fact that $\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2)$ vanishes for only half the electrons in the exchange case. Thus the effect of adding correlation is to deepen and narrow the hole. Indeed this picture works almost quantitatively. For example for Na with $r_s = 4.0$

the exchange energy alone is $-.229\,0$ Ry, while the sum of exchange and correlation is $-.299\,6$ Ry. Using a picture of a flat hole with twice the depth in the exchange plus correlation case, we should expect correlation to increase the energy by a factor of $2^{\frac{1}{3}}$ which gives $-.289\,0$. The two holes are sketched in Fig. 1.

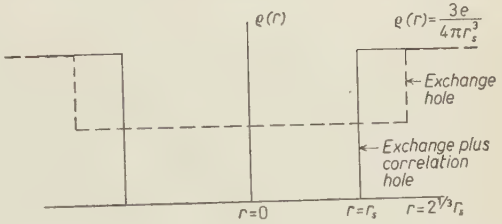


Fig. 1. – Distribution of charge density for exchange hole and exchange plus correlation hole.

It is interesting to note that in cases where we have less than one valence electron per atom the correlation hole may actually become larger than the unit cell of the crystal. In this case we should expect that the total interaction energy would become negative. Physically this arises from the fact that the hole will extend into neighboring cells and expose the electron in a given cell to some of the positive ionic charge in neighboring cells.

At the surface of a metal the correlation hole goes over into the image charge of classical electrostatic theory. In effect the hole broadens in extent and hangs behind the electron in regions of high charge density. Such an effect will occur also in the interior of a metal when there are fluctuations in electronic charge density. The hole will always tend to lag behind the electron in regions of high charge accumulation, and will thus favor more pile-up of charge in such regions. It is probably an effect such as this which is responsible in part for the transition from the Bloch approximation to the Heitler-London approximation. The Bloch approximation works best when the charge density is fairly uniform and the hole therefore modifies the energy of the electron without having much influence on its spatial distribution. We shall come back to this point in the next section.

4. – The effective one-electron potential.

We may consider the various energy terms in (2.14) and (2.17) in the light of an effective potential for the electrons. We may write such an effective potential in the form:

$$(2.25)\qquad V(\boldsymbol{x}_1) = -\frac{2Z}{r_{10}} + 2\int \frac{\gamma_0(\boldsymbol{x}_2)}{r_{12}}\,\mathrm{d}\boldsymbol{x}_2 + 2\sum_{g\neq 0}\int \frac{\gamma_{g,n}(\boldsymbol{x}_2)}{r_{12}}\,\mathrm{d}\boldsymbol{x}_2 + \\ + 2\int \frac{(2\Gamma(\boldsymbol{x}_1,\boldsymbol{x}_2)/\Gamma(\boldsymbol{x}_1)) - \Gamma(\boldsymbol{x}_1)}{r_{12}}\,\mathrm{d}\boldsymbol{x}_2\,,$$

where we have considered the potential in the 0'th cell, and where we have separated out the non-spherical part of the charge distribution

$$\gamma_g(\boldsymbol{x}_1) = \gamma_{g,s}(\boldsymbol{x}_1) + \gamma_{g,n}(\boldsymbol{x}_1) \,. \tag{2.26}$$

The subscript s denoting spherical symmetry, and the subscript n the non-spherically symmetric part. With the aid of (2.25) we may list the various contributions to the potential seen by an electron in cell 0 as follows:

1) The electrostatic potential arising from the nucleus in cell 0. This is the first term of (2.25).

2) The electrostatic potential arising from the core electrons in cell 0. This is part of the second term in (2.25) if we assume that we can separate $\gamma_0(\boldsymbol{x}_1)$ into a core and valence part.

3) The electrostatic potential arising from the other valence electrons within cell 0. This will be the other part of the second term in (2.25).

4) The electrostatic potential due to the non-spherically symmetric part of the electronic charge distribution in all cells other than 0. This is the third term of (2.25).

5) The exchange effect between the core and valence electrons.

6) The exchange effect arising from all other valence electrons of the same spin.

7) The correlation or polarization effect arising from the correlation between the valence electron and core electron motions.

8) The correlation effect between valence electrons. All the effects 5, 6, 7, and 8 are included in the last term of (2.25).

The electrostatic terms are easily visualized and require no further discussion. It need only be noted that it is only the non-spherical part of the charge distribution in other cells which contributes to the potential. This is a small effect and is ignored in most calculations. It should be remarked in passing, however, that the changes in energy resulting from small distortions of the crystal which do not change the volume arise almost entirely from these terms in many cases. Thus, although such terms are negligible from the standpoint of cohesion, they may make the major contribution to the elastic constants, as is the case in the alkali metals [12].

If we omit the last term of (2.25) we have the potential for the Hartree approximation. In the Hartree method of calculation we determine one electron wave function s from the solution of the wave equation in this po-

tential. With the resulting wave functions we may determine charge densities and recompute a new potential. This procedure is then repeated until it is self-consistent, i.e. the potential determined for the wave functions agrees with the potential for which they are solutions of the Schrödinger equation.

If we retain the last term of (2.25) and assume that the wave functions are Slater determinants of one-electron functions, then we obtain a new form of self-consistent problem. This is, in fact, a modification of the Hartree-Fock problem suggested by SLATER [6]. It includes a contribution to the potential arising from the exchange hole, but this hole is taken the same for every state of the electron. Thus in this approximation every electron moves in the same potential. In a more rigorous formulation of the problem it turns out that the form of the exchange hole should be taken as different for each state of the electron. The Slater approximation represented by (2.25) is adequate for handling exchange between valence electrons, although even here the exchange potential for free electrons varies by a factor of 2 from the bottom to the top of the Fermi distribution. For handling both exchange and correlation effects between valence electrons and core, however, it is found necessary to set up a different potential for each angular momentum of an electron. In other words the radial part of the wave function of a valence electron must satisfy a differential equation

$$\frac{\mathrm{d}^2 f_0}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}f_0}{\mathrm{d}r} + \left[E - V_l(r) - \frac{l(l+1)}{r^2}\right] f_0 = 0\,, \tag{2.27}$$

where $V_l(r)$ is different for each value of l. A general solution of the Schrödinger equation then takes the form

$$\Psi = \sum_{l,m} A_{l,m} f_l(r) Y_{l,m}(\theta, \varphi)\,, \tag{2.28}$$

where the $Y_{l,m}(\theta, \varphi)$ are spherical harmonics, and the $A_{l,m}$ are expansion coefficients. The adequacy of this representation of exchange and correlation effects of core and valence electrons has been discussed by HAM [7] and by CALLAWAY [8]. It is best justified in cases where the characteristic energies of the core and valence electrons differ most widely.

When the electron is outside the ion core, the electrostatic potential due to the ion core may be written $-2v/r$, where v is the valence. The exchange contribution to the potential can be shown to vanish outside the ion core because it is proportional to the amplitude of the core wave functions. The correlation or polarization potential, which may be very complicated inside the ion core, takes the simple form $-2\alpha'/r^4$, where α' is the polarizability of the ion core measured in atomic units. Although this contribution to the

potential is small, it has a significant effect in calculations for the heavier metals and must be taken into account in any quantitative calculation as we shall see. Its greatest influence on the energy arises from the potential inside the ion core where its form is not known at all. Polarization potentials determined semi-empirically have been used in calculation for solids by HEINE [9] and CALLAWAY [8]. The justification for representing the correlation between core and valence electrons has been discussed by HAM [7]. It is only justified when the core electrons may be regarded as responding essentially adiabatically to the much slower motion of the valence electron.

We have already mentioned that the correlation hole tends to move along with the valence electron in so far as the correlation and exchange between valence electrons is concerned. In fact for plane waves this is exactly true and in consequence the effective potential due to exchange and correlation effects may be regarded as a constant. In many real crystals we shall see that the charge density due to the valence electrons is constant over most of the cell. In view of this, considering that the exchange and correlation hole is of the order of the size of the cell, it seems likely that in simple metals it is a good approximation to regard the exchange and correlation potential as simply a constant, which may be of different magnitude for each state of the valence electron. It has sometimes been argued that because of the smallness of the interaction energy for monovalent metals we may consider that exchange and correlation just succeed in keeping all other electrons out of a given cell, so that we should take as the potential seen by the valence electron in the solid just the potential of the bare ion core [10]. This procedure is highly questionable, however, because the correlation hole moves with the electron while the contribution to the electrostatic potential from the other valence electrons remains fixed in the crystal. This point has been especially emphasized by HEINE [9]. Thus whereas the constant potential may simply be incorporated in the eigenvalue, the electrostatic potential should be taken into account explicitly in computing the wave functions. This point will be taken up again at the end of Chapter V. We thus end up with a contribution to the potential from the other valence electrons which is given by:

$$V(r) = \frac{3}{r_s} - \frac{r^2}{r_s^3}. \tag{2.29}$$

This potential cancels exactly against the ion core potential at the point $r = r_s$, which is the radius of the spherical unit cell approximating to the real cell, i.e. the Wigner-Seitz sphere.

The justification for representing exchange and correlation effects between valence electrons as a one-electron potential will be discussed further by Professor PINES [11].

REFERENCES

[1] P. O. LOWDIN: *Phil. Mag. Supp.*, **5**, 1 (1956); cfr. especially Sect. **2.1**, pp. 5-20.

[2] F. SEITZ: *Modern Theory of Solids* (New York, 1940), chap. II; E. MADELUNG: *Gött. Nach.*, 1909, p. 100 and 1910, p. 43; *Phys. Zeits.*, **11**, 898 (1910): P. P. EWALD: *Thesis* (Munich, 1912).

[3] F. SEITZ: *Modern Theory of Solids* (New York, 1940), p. 364.

[4] F. SEITZ: *Modern Theory of Solids* (New York, 1940), pp. 339-342; ref. (1), p. 9; J. C. SLATER: *Rev. Mod. Phys.*, **6**, 209 (1934).

[5] D. PINES: *Solid State Physics*, vol. I, pp. 368-449, cfr. in particular eq. (6.24) p. 397, and the discussion of Sect. **6**, pp. 391-400.

[6] J. C. SLATER: *Phys. Rev.*, **81**, 385 (1951).

[7] F. S. HAM: *Solid State Physics*, vol. I, pp. 127-192, cfr. especially Sect. **9**, pp. 169-174.

[8] Exchange: J. CALLAWAY: *Phys. Rev.*, **103**, 1219 (1956); S. BERMAN, J. CALLAWAY and R. D. WOODS: *Phys. Rev.*, **101**, 1467 (1956). Correlation: J. CALLAWAY: Technical Report No. 3, Feb. 15, 1957 (University of Miami, unpublished).

[9] V. HEINE: *Proc. Roy. Soc.*, **240**A, 340 (1957); *Thesis* (Camb. U., 1956, unpublished).

[10] F. SEITZ: *Modern Theory of Solids*, pp. 355-356.

[11] For some orientation on this question see also D. PINES: *Solid State Physics*, vol. I, Sect. **3**˙8.

[12] F. SEITZ, ref. [2].

III.

Solution of the one electron problem: energy bands.

1. – Introduction.

It is not our aim to give in this chapter a comprehensive review of the theory of energy bands, but merely to supply enough of the general theory to provide a sound basis for the discussion of the method of Wigner and Seitz, and a critical evaluation of its limitations.

We have seen that one of the basic problems is the solution of the Schrödinger equation for the motion of a single electron in a perfectly periodic potential, i.e. one which satisfies the condition:

$$\begin{cases} V(\boldsymbol{r} + \boldsymbol{R}) = V(\boldsymbol{r}) , \\ \\ \boldsymbol{R} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3 . \end{cases} \tag{3.1}$$

Here $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ are the three primitive translation vectors of the crystal lattice, n_1, n_2, n_3 are integers. We shall see later that it is actually not necessary that $V(\boldsymbol{r})$ be a simple potential, but it can actually include exchange and correlation effects to some extent. We shall stay with the simpler form in the present derivations, however.

It is proved in many standard texts that the most general normalizable solutions of the Schrödinger equation with a potential satisfying (3.1) satisfy the Bloch condition

$$b_n(\boldsymbol{k}, \boldsymbol{r}+\boldsymbol{R}) = \exp[i\boldsymbol{k}\cdot\boldsymbol{R}]b_n(\boldsymbol{k}, \boldsymbol{r}) . \tag{3.2}$$

Here $\boldsymbol{k}$ is the reduced wave vector, to be defined later. (3.2) is an expression of the translational symmetry of the crystal. Associated with the wave function which satisfies (3.2) is an energy eigenvalue $E_n(\boldsymbol{k})$. For a fixed choice of $\boldsymbol{k}$ we find only a discrete set of energies associated with normalizable wave functions. These are labelled by n, which may be taken to be the number of the level in order of increasing energy.

Suppose we increase $\boldsymbol{k}$ by a certain vector $\boldsymbol{K}$, such that $\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{K}$. Then there exists a set of vectors $\boldsymbol{K}$ which have the property

$$\exp[i\boldsymbol{K}\cdot\boldsymbol{R}] = 1 . \tag{3.3}$$

and the corresponding set of $\boldsymbol{K}$ are the vectors of the reciprocal lattice. The reciprocal lattice is constructed according to the following rules. We construct a set of vectors $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$ which are reciprocal to the primitive translation vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, that is, which satisfy the condition:

$$\boldsymbol{a}_i\cdot\boldsymbol{b}_j = \delta_{ij} . \tag{3.4}$$

Out of the $\boldsymbol{b}$ vectors we construct a lattice defined by:

$$\boldsymbol{K} = 2\pi(h_1\boldsymbol{b}_1 + h_2\boldsymbol{b}_2 + h_3\boldsymbol{b}_3) , \tag{3.5}$$

where h_1, h_2, h_3 are integers. It is readily verified that any vector defined by (3.5) satisfies the condition (3.3) for arbitrary $\boldsymbol{R}$. The totality of points defined by all the vectors $\boldsymbol{K}$ drawn from a common origin constitutes a lattice, which is the reciprocal lattice, and the vectors are called the reciprocal lattice vectors.

Reciprocal lattice for a given real lattice is unique if the primitive lattice vectors are chosen so as to span the real lattice in the simplest way, i.e. to generate a Bravais lattice [19]. For example, for bcc the primitive vectors

are taken as

$$\boldsymbol{a}_1 = \tfrac{1}{2}a_0(-\hat{x} + \hat{y} + \hat{z}), \qquad \boldsymbol{a}_2 = \tfrac{1}{2}a_0(\hat{x} - \hat{y} + \hat{z}), \qquad \boldsymbol{a}_3 = \tfrac{1}{2}a_0(\hat{x} + \hat{y} - \hat{z}) \tag{3.6}$$

where a_0 is the cube edge, and $\hat{x}$, $\hat{y}$, $\hat{z}$ are unit vectors in the cartesian coordinate directions. The reciprocal vectors are

$$\begin{cases} \boldsymbol{b}_1 = \dfrac{\boldsymbol{a}_2 \times \boldsymbol{a}_3}{\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 \times \boldsymbol{a}_3} = \dfrac{1}{a_0}(\hat{y} + \hat{z}) \\ \boldsymbol{b}_2 = \dfrac{\boldsymbol{a}_3 \times \boldsymbol{a}_1}{\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 \times \boldsymbol{a}_3} = \dfrac{1}{a_0}(\hat{x} + \hat{z}) \\ \boldsymbol{b}_3 = \dfrac{\boldsymbol{a}_1 \times \boldsymbol{a}_2}{\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 \times \boldsymbol{a}_3} = \dfrac{1}{a_0}(\hat{x} + \hat{y}) \end{cases} \tag{3.7}$$

and it follows that the reciprocal lattice is fcc with the cube edge $4\pi/a_0$. Similarly, the lattice reciprocal to fcc is body-centered with a cube edge $4\pi/a_0$.

$\boldsymbol{k}$-vectors which differ by a vector of the reciprocal lattice cannot be physically distinguished, because the corresponding wave functions have indistinguishable translational symmetries according to (3.2). We may either regard $b_n(\boldsymbol{k}, \boldsymbol{r})$ and $E_n(\boldsymbol{k})$ as functions which have translational periodicity in $\boldsymbol{k}$-space, i.e. which satisfy the conditions:

$$\begin{cases} b_n(\boldsymbol{k} + \boldsymbol{K}, \boldsymbol{r}) = b_n(\boldsymbol{k}, \boldsymbol{r}), \\ E_n(\boldsymbol{k} + \boldsymbol{K}) = E_n(\boldsymbol{k}), \end{cases} \tag{3.8}$$

or, alternatively, we may restrict the vector $\boldsymbol{k}$ to a region of $\boldsymbol{k}$-space such that no two points in this region are separated by any vector $\boldsymbol{K}$. This is a unit cell in reciprocal space, often referred to as the first Brillouin Zone, or 1st BZ. Like the atomic polyhedron in the real lattice (if it is a Bravais lattice having one atom per unit cell), it may be constructed by choosing a particular lattice point and enclosing it with the smallest polyhedron formed by constructing the perpendicular bisector planes of all the vectors $\boldsymbol{R}$ or $\boldsymbol{K}$ drawn from the origin to all the points of the lattice or reciprocal lattice, respectively. The restricted domain of $\boldsymbol{k}$-vectors is known as the reduced wave vectors.

Discussion of symmetry properties is outside the scope of this review. The reader is referred to the literature [2]. We may note a few salient points, however. Reciprocal lattice is always a Bravais lattice. In general it has the same point group as the real lattice, except that its symmetry group always includes the inversion. Thus, the reciprocal to the diamond structure, which is not invariant under inversion, is a bcc reciprocal lattice which has full cubic point symmetry. If we label energy levels for all $\boldsymbol{k}$ always in order

of increasing or decreasing energy, we may define a set of surfaces of constant energy in $\boldsymbol{k}$ space by $E_n(\boldsymbol{k}) = E$. These energy surfaces have the same point symmetry as the reciprocal lattice. In particular

$$E_n(\boldsymbol{k}) = E_n(-\boldsymbol{k}) \,. \tag{3.9}$$

If we neglect magnetic interactions, i.e. if we have a real hamiltonian, then we also have

$$b_n(\boldsymbol{k}, \boldsymbol{r}) = b_n^*(-\boldsymbol{k}, \boldsymbol{r}) \,. \tag{3.10}$$

In the presence of magnetic interactions the situation is more complicated [3].

2. – The cellular method. General theory.

The Bloch condition (3.1) can be reformulated into a boundary condition (linear and homogeneous) to be satisfied by the wave function entirely within one unit cell of the crystal. It is customary to define the unit cell in terms of a symmetrical polyhedron surrounding some atom and constructed by means of perpendicular bisector planes as defined previously. The unit cell has the property that points on opposite faces of it differ in position by some vector $\boldsymbol{R}$ which is a translation vector as defined in (3.1). Thus by means of (3.2) a wave function on one face of the cell may be expressed in terms of a wave function on the other face. The condition that the function be continuous with continuous derivatives throughout the crystal then reduces to the following relations:

$$\begin{cases} \exp\left[-i\boldsymbol{k}\cdot\boldsymbol{r}_s'\right] b_n(\boldsymbol{k}, \boldsymbol{r}_s') = \exp\left[-i\boldsymbol{k}\cdot\boldsymbol{r}_s\right] b_n(\boldsymbol{k}, \boldsymbol{r}_s) \\ -\exp\left[-i\boldsymbol{k}\cdot\boldsymbol{r}_s'\right]\hat{n}\cdot\nabla b_n(\boldsymbol{k}, \boldsymbol{r}_s') = \exp\left[-i\boldsymbol{k}\cdot\boldsymbol{r}_s\right]\hat{n}\cdot\nabla b_n(\boldsymbol{k}, \boldsymbol{r}_s) \,, \end{cases} \tag{3.11}$$

where

$$\boldsymbol{r}_s' - \boldsymbol{r}_s = \boldsymbol{R} \tag{3.12}$$

and where $\boldsymbol{r}_s$, $\boldsymbol{r}_s'$ are two points on the surface of the cell which are related by (3.12), the right hand side representing primitive translation vectors. (3.11) is valid for any definition of unit cell, but it is not very useful unless the cell is chosen as the symmetrical polyhedron as described above. The reason for this is that inside this polyhedron the Schrödinger equation is nearly spherically symmetric so that differential eq. is approximately separable.

Another way of writing (3.11) which is equivalent. The Bloch condition (3.2) is also equivalent to

$$b_n(\boldsymbol{k}, \boldsymbol{r}) = \exp\left[i\boldsymbol{k}\cdot\boldsymbol{r}\right] u_n(\boldsymbol{k}, \boldsymbol{r}) \,, \tag{3.13}$$

where $u_n(\boldsymbol{k}, 0)$ is a periodic function of $\boldsymbol{r}$, and satisfies a boundary condition:

$$u_n(\boldsymbol{k}, \boldsymbol{r}_s') = u_n(\boldsymbol{k}, \boldsymbol{r}_s) - \hat{n}' \cdot \nabla u_n(\boldsymbol{k}, \boldsymbol{r}_s') = \hat{n} \cdot \nabla u_n(\boldsymbol{k}, \boldsymbol{r}_s) \,, \tag{3.14}$$

where in both (3.11) and (3.14) the operators $\hat{n} \cdot \nabla$ and $\hat{n}' \cdot \nabla$ are normal derivatives, taken normal outwards to the face of the cell at $\boldsymbol{r}_s$ and $\boldsymbol{r}_s'$, respectively.

Suppose the potential $V(\boldsymbol{r})$ is spherically symmetric throughout the atomic polyhedron. The wave function satisfies the spherically symmetric Schrödinger equation

$$\nabla^2 b_n(\boldsymbol{k}, \boldsymbol{r}) + [E - V(r)] b_n(\boldsymbol{k}, \boldsymbol{r}) = 0 \,, \tag{3.15}$$

where Rydberg units are used for energy and Bohr units for length. The general solution of (3.15) may be expanded in spherical harmonics in the form

$$b_n(\boldsymbol{k}, \boldsymbol{r}) = \sum_{l,m} A_{l,m} f_l(E, r) Y_{l,m}(\theta, \varphi) \,, \tag{3.16}$$

where the $A_{l,m}$ are arbitrary coefficients, and the radial functions satisfy the radial differential equations:

$$\frac{\mathrm{d}^2 f_l}{\mathrm{d}r^2} + \frac{2}{r} \frac{\mathrm{d}f_l}{\mathrm{d}r} + \left[E - V(r) - \frac{l(l+1)}{r^2} \right] f_l = 0 \,. \tag{3.17}$$

In (3.16) it is assumed that the f's are normalized in such a way that

$$f_l(E, r_s) = 1 \tag{3.18}$$

for some choice of r_s and for all E.

If we substitute the expansion (3.16) into the boundary condition (B.C.) (3.11), we obtain in principle an infinite system of linear homogeneous algebraic equations for the determination of the A's. The compatibility condition for this set leads to a single relation between the f's, corresponding to the vanishing of the determinant of the linear equations. This condition is in general only satisfied for a discrete infinite set of E's, which are the eigenvalues $E_n(\boldsymbol{k})$.

In practice, of course, the boundary value problem is only tractable if we cut-off the expansion (3.16) after a finite number of terms. This means that we must also use only a finite number of boundary conditions, and the trick then becomes to select these in the optimum way.

There are some tricky points to be noticed. Any Bloch function may be expanded in the form (3.6) but the expansion coefficients only satisfy the differential equations (3.17) for values of r such that the potential is spherically symmetric on the entire sphere. This condition is not usually satisfied for

spheres larger than the inscribed sphere for the atomic polyhedron. Larger spheres overlap into neighboring polyhedra and the potential becomes that which is centered on that polyhedron. From the standpoint of the given polyhedron, the potential is no longer spherically symmetric. The only exception to this difficulty occurs if the potential is taken as constant outside the inscribed sphere. One practice is to arbitrarily choose the potential in this way, satisfy the boundary conditions, and then treat any deviations from constant potential as a perturbation of the solution. When a non-constant potential is used outside the inscribed sphere, most frequently a $1/r$ potential, and when the coefficients in (3.16) are treated as solutions of the radial equation, the solutions become somewhat oversensitive to the exact choice of boundary conditions. This is not a prohibitive situation, but makes for some practical difficulties in applying the cellular method [20].

3. – Brief history of the cellular method.

It is not the primary purpose of this review to go into band calculations, and so I will only sketch briefly how the cellular method has been used, and will make very little reference to other methods [3].

The original application was suggested by SLATER [4]. He proposed that the boundary conditions be satisfied at selected points on the cell boundary, the number of non-equivalent points being chosen to just determine the number of unknown coefficients retained in (3.16). The general method was applied by other authors somewhat indiscriminately to a number of solids before Shockley [5] investigated its accuracy by a rather neat test problem, the so-called empty lattice test. We know that in a constant potential the Bloch functions are plane waves while the solutions of (3.17) are spherical Bessel functions. We may then use these Bessel functions in the approximate B.C., determine the corresponding $E_n(\boldsymbol{k})$, and compare them with the known correct answer, which is just

$$E_n(\boldsymbol{k}) = (\boldsymbol{k} + \boldsymbol{K}_n)^2 , \tag{3.19}$$

where $\boldsymbol{K}_n$ is a reciprocal lattice vector. Shockley found that for the boundary conditions at selected points that had actually been used in practical calculations the empty lattice test was extremely poor in some cases. In general the results were poorer the higher the energy above the lowest state of the valence band. Unfortunately the main conclusion from Shockley's work was that in order to obtain reasonable accuracy for excited states it was necessary to retain so many terms in the expansion (3.16) and satisfy boundary conditions at so many points that for general values of $\boldsymbol{k}$ the procedure became impractical.

The first attack on this impasse was proposed by BETHE and VON DER LAGE [6]. They emphasized that for certain special values of $\boldsymbol{k}$, many terms in (3.16) could be omitted by symmetry arguments. For such cases it is possible to retain harmonics of high order in the expansion (3.16) while still keeping the total number of harmonics manageable. For example, in the case of Na for the points $\boldsymbol{k} = (0, 0, 0)$ or $(0, 0, 4\pi/a_0)$) it is possible to retain all harmonics up to order $l = 8$ while still having only 3 harmonics in the expansion (3.16). This is true for functions of s-like symmetry. BETHE and VON DER LAGE applied this idea to Na, using B.C. which were a mixture of conditions at specific points on the cell surface and averages over certain lines.

Quite similar procedures were suggested by HOWARTH and JONES [7]. They used the same expansions as BETHE and VON DER LAGE but satisfied the B.C. in a different way. Essentially they used a redundant number of boundary points, averaging their results over different sets of points. The points themselves were selected on the basis of their capacity to give good results with the empty lattice test. HOWARTH and JONES also developed a procedure for correcting their final eigenvalues by means of a boundary perturbation integral. This integral will play an important role in the methods to be discussed below, and we shall postpone its detailed discussion until that time.

The next development in the cellular approach was provided by W. KOHN [12] who showed that the correction perturbation integral suggested by HOWARTH and JONES [7] could be used as the basis of a variational principle for the satisfaction of the B.C. Again the Bloch function was to be expanded in the form (3.16) and the coefficients determined by making stationary a certain surface integral over the atomic polyhedron. Practical applications of this method have been made by PINCHERLE and others [13].

The basic philosophy of the cellular method is to expand the Bloch function in terms of exact solutions of the Schrödinger equation, and to determine the unknown coefficients in this expansion by attempting to satisfy the cellular boundary conditions in an approximate way. As might be expected the converse approach is also possible, namely to expand the Bloch function in terms of functions which satisfy the B.C. exactly but which satisfy the differential equation only approximately. The unknown coefficients in the expansion are then determined by satisfying the Schrödinger equation to a closer and closer approximation. The best example of this procedure is the method of orthogonalized plane waves, originally suggested by HERRING [8] and exploited extensively by HERMAN [9], CALLAWAY [10], PARMENTER [11] and others. As the name implies, in this method, the expansion functions are taken to be plane waves of the form

$$\exp\left[i(\boldsymbol{k} + \boldsymbol{K}_n)\cdot\boldsymbol{r}\right], \tag{3.20}$$

where the different reciprocal lattice vectors $\boldsymbol{K}_n$ are associated with different

waves in the expansion. The waves (3.20) are modified so as to be made orthogonal, by the Schmidt method, to the wave functions of all the ion core electrons. The trick here is that the wave functions of the ion core electrons are highly localized and in fact vanish for all practical purposes on the atomic polyhedron. Thus the process of orthogonalizing the waves (3.20) to the core electrons does not modify their form in the immediate vicinity of the cell boundaries. In consequence the orthogonalized plane waves still satisfy the exact boundary conditions of the problem, but are much better approximations to solutions of the Schrödinger equation than plane waves alone.

We shall not review the several other methods of calculation which have been proposed and tried in practice, since they will not enter into our discussion of cohesive energies.

4. – The sphere approximation.

WIGNER and SEITZ first made the suggestion [14] of replacing the exact boundary conditions (3.11) or (3.14) by a spherical approximation. The idea was to replace the polyhedral unit cell by a sphere of equal volume. In the B.C. $\boldsymbol{r}_s$ and $\boldsymbol{r}_s'$ become radius vectors from the center of the sphere to diametrically opposite points on the surface. Since the potential is also assumed spherically symmetric, there is no preferred axis in the boundary value problem other than the direction of the vector $\boldsymbol{k}$. Thus the expansion (3.16) reduces to the simpler form

$$b_n(\boldsymbol{k}, \boldsymbol{r}) = \sum_{l=0}^{\infty} A_l f_l(E, r) P_l(\hat{k} \cdot \hat{r}) , \tag{3.21}$$

where P_l is a Legendre polynomial of order l, $\hat{k}$ and $\hat{r}$ represent unit vectors, so that $\hat{k} \cdot \hat{r}$ is the cosine of the angle between $\boldsymbol{k}$ and $\boldsymbol{r}_s$.

Taking into account the oddness and eveness of the Legendre polynomials, the boundary conditions (3.11) may be written in the form:

$$\begin{cases} \sum_{l=0}^{\infty} A_l f_l(E, r_s) P_l(\cos\theta)(\exp[-ikr_s\cos\theta] - (-1)^l \exp[ikr_s\cos\theta]) = 0 , \\ \sum_{l=0}^{\infty} A_l f_l'(E, r_s) P_l(\cos\theta)(\exp[-ikr_s\cos\theta] + (-1)^l \exp[ikr_s\cos\theta]) = 0 , \end{cases} \tag{3.22}$$

where $\cos\theta = \hat{k} \cdot \hat{r}$, and where the prime in the second equation indicates differentiation with respect to r_s, the radius of the spherical cell. One way of solving the system (3.22) is to multiply each equation by $P_L(\cos\theta)$ and in-

tegrate over all directions of $\boldsymbol{r}$. The resulting equations are:

$$\text{(3.23)} \qquad \text{Even } L: \quad \sum_{l=0}^{\infty} A_l f_l'(E, r_s) \int_{-1}^{+1} \exp[-ikr_s\mu] P_l(\mu) P_L(\mu)\, d\mu = 0 ,$$

$$\text{(3.24)} \qquad \text{Odd } L: \quad \sum_{l=0}^{\infty} A_l f_l(E, r_s) \int_{-1}^{+1} \exp[-ikr_s\mu] P_l(\mu) P_L(\mu)\, d\mu = 0 .$$

It is to be noted that for a given value of L one of the two equations (3.22) is always satisfied identically, so that half of the system is always redundant. In practice, of course, the equations are solved by retaining only a finite number of harmonics. In order to make the system just compatible we retain the same number of L's as l's, in which case it is easily seen that the number of equations equals the number of unknowns. The energy is obtained from the condition that the determinant of the system vanish.

It is of interest to apply the empty lattice test to the set (3.23)–(3.24). We first note that the equations for the two lowest values of L, i.e. $L = 0$ and $L = 1$, take the form:

$$\text{(3.25)} \qquad \sum_{l=0}^{\infty} (-1)^l A_l f_l'(E, r_s) j_l(kr_s) = 0 , \qquad \sum_{l=0}^{\infty} (-1)^l A_l f_l(E, r_s) j_l'(kr_s) = 0 ,$$

where $j_l(kr_s)$ is a spherical Bessel function and $j_l'(kr_s)$ is its first derivative with respect to its argument. If we expand the determinant of (3.23)–(3.24) by minors of the first two rows, eq. (3.25) shows us that each term in the expansion contains a factor of the form

$$\text{(3.26)} \qquad \{f_l'(E, r_s) f_m(E, r_s) j_l(kr_s) j_m'(kr_s) - f_m'(E, r_s) f_l(E, r_s) j_m(kr_s) j_l'(kr_s)\} .$$

For the empty lattice we have

$$\text{(3.27)} \qquad f_l(E, r_s) = j_l(\sqrt{E}\, r_s) ,$$

where it is assumed that E is measured from the zero of kinetic energy. From (3.26) we see that if we take $E = k^2$ then it follows that each minor in the expansion of the determinant vanishes identically, and therefore this value of E is an eigenvalue. If we cut off the expansion after a finite number of harmonics there will, of course, be other allowed values of E. However, it seems probable, though not proved, that as we allow the secular determinant to go to infinite order the spurious roots go off towards positive infinity, and we are left only with the correct empty lattice energy $E = k^2$. The ability

to fulfill the empty lattice test when the harmonic expansion is cut off at finite order is peculiar to the particular form of the spherical boundary conditions taken in (3.23)–(3.24). There are other ways of writing the conditions, however. For example, we could have multiplied each equation in (3.22) by $\exp[-ikr_s\cos\theta]$ before multiplying by P_L and integrating. We would then have obtained a system which is equally valid and entirely equivalent to (3.23)–(3.24) when carried to infinite order, but different when cut off at finite order. Such a system would satisfy the empty lattice test only in the limit of infinite order, but the finite system would not do so. It is for this reason that we prefer the particular form, since in a real calculation it insures that the secular determinant will nearly vanish if the radial functions are almost like Bessel functions, as is in fact the case for many simple solids, as we shall see later. The use of the special set (3.23)–(3.24) gives us the greatest accuracy in the energy for a given number of harmonics.

We note also in connection with these equations that there is no reason why we need restrict ourselves to reduced values of the wave vector $\boldsymbol{k}$, since the empty lattice test is satisfied for all values of $\boldsymbol{k}$. That values of $\boldsymbol{k}$ outside the 1st BZ are significant simply follows from the fact that the artificial spherical boundary conditions we have used are not compatible with the periodicity of the crystal. Nevertheless these apparently artificial solutions will prove quite useful when we come to discuss the exact boundary conditions.

5. – The Wigner-Seitz solution.

We consider first the case $\boldsymbol{k}=0$. The B.C. (3.23)–(3.24) reduce to the simple form:

(3.28a) Even L: $A_L f'_L(E, r_s) = 0$,

(3.28b) Odd L: $A_L f_L(E, r_s) = 0$.

Only one non-vanishing harmonic occurs in the expansion, and if the non-vanishing harmonic is even, the radial derivative of the radial function vanishes on the boundary, while if the non-vanishing harmonic is odd, it is the function itself that vanishes.

In the case of Na Wigner and Seitz chose $L=0$, since the lowest valence state of the free atom is an s-state. The value of E for which (3.28a) is satisfied when $L=0$ then represents the energy for $\boldsymbol{k}=0$ in the valence band of Na in the solid. This solution satisfies the empty lattice test, since

(3.29) $$j'_0(\sqrt{E}\, r_s) = 0\,,$$

when $E=0$.

We could attempt to use (3.28) to obtain higher states, but we run into difficulties when we try to apply the empty lattice test in such cases. As an illustration, let us consider the next lowest set of free electron states in the BCC lattice corresponding to a reduced wave vector of 0. According to (3.19) this energy level is at K^2_{110} where K_{110} is the length of the shortest reciprocal lattice vector, (R.L.V.) which is $2\sqrt{2}\,\pi/a_0$. There are 12 such R.L.V. of the same length, and the corresponding energy level is 12-fold degenerate. From group theory considerations it is possible to construct linear combinations which have s, p, d, and f-like symmetry. There is 1 s combination, 3 p combinations, 5 d combinations, and 3 f combinations, all of which have the same energy in an empty lattice. In a real crystal these combinations will have 5 different energies, since the d combinations give rise to two different energies which are triply and doubly degenerate in cubic symmetry; but this need not concern us for the moment. Table XI shows the positions of these various levels for

TABLE XI.

	Sphere	Free	Na-extended	True Na
s	1.265	1.210	1.224	
p	1.265	1.210	1.224	1.19
d	0.70	1.210	1.224	1.16 (*)
f	3.06	1.210	1.224	1.23 (*)

(*) Na values measured from E_0.

$r_s = 4.0$ in the empty lattice sphere approximation. The correct result is given for comparison. (The table illustrates two points, 1st that the sphere approximation is poor except when $\boldsymbol{k}$ is defined in the extended sense, and 2nd that the values of the radial wave functions are very close to the free electron values for Na.) The 3rd column shows the result obtained for the sphere approximation in Na when the spherical B.C. are applied with $\boldsymbol{k}$ defined in the extended sense. In this case the energies are measured from the bottom of the valence band.

6. – Energy away from $\boldsymbol{k} = 0$ in the sphere approximation.

BARDEEN first suggested a method [15] for computing the energy away from $\boldsymbol{k} = 0$, and his method has been extended by SILVERMAN [16] and by KOHN [17]. These arguments are rather complicated, and we shall use instead a different argument based directly on the B.C. (3.23)–(3.24). The object of

the calculation is to obtain the energy as a power series in k^2:

$$E = E_0 + E_2 k^2 + E_4 k^4 \ldots , \tag{3.30}$$

E_0 is the solution of (3.28a), E_2 is given by the Bardeen method, and E_4 is given by the extension due to SILVERMAN.

We begin by writing (3.23) and (3.24) explicitly, assuming that $f(E, r_s) = 1$ for all E. The resulting system of equations, up to the 4th order in l and L is:

$$\begin{vmatrix} j_0\varphi_0 & -ij_1\varphi_1 & -j_2\varphi_2 & ij_3\varphi_3 \\ -ij_1 & \frac{1}{3}j_0 - \frac{2}{3}j_2 & -i\left(\frac{2}{5}j_1 - \frac{3}{5}j_3\right) & -\left(\frac{3}{7}j_2 - \frac{4}{7}j_4\right) \\ -j_2\varphi_0 & -i\left(\frac{2}{5}j_1 - \frac{3}{5}j_3\right)\varphi_1 & \left(\frac{1}{5}j_0 - \frac{2}{7}j_2 + \frac{18}{35}j_4\right)\varphi_2 & -i\left(\frac{9}{35}j_1 - \frac{4}{15}j_3 + \frac{10}{21}j_5\right)\varphi_3 \\ ij_3 & -\left(\frac{3}{4}j_2 - \frac{4}{7}j_4\right) & -i\left(\frac{9}{35}j_1 - \frac{4}{15}j_3 + \frac{10}{21}j_5\right) & \left(\frac{1}{7}j_0 - \frac{4}{21}j_2 + \frac{18}{77}j_4 - \frac{100}{231}j_6\right) \end{vmatrix} \tag{3.31}$$

Here we have written down the matrix of the system. The successive columns refer to the coefficients A_0, A_1, A_2 and A_3, respectively. The j's are spherical Bessel functions with the arguments kr_s. The φ_l's are defined by

$$\varphi_l = \frac{r_s(f_l'(E, r_s)}{f_l(E, r_s)} . \tag{3.32}$$

We now expand the system (3.31) in powers of k. In doing this we also assume the expansion (3.30) and use a Taylor's expansion for the φ's in energy about the ground state energy E_0. Thus, for example, we take

$$\varphi_0 \to \varphi_0 + k^2\left[E_2\frac{\partial\varphi_0}{\partial E}\right] + k^4\left[E_4\frac{\partial\varphi_0}{\partial E} + \frac{1}{2}E_2\frac{\partial^2\varphi_0}{\partial E^2}\right], \tag{3.33a}$$

$$\varphi_1 \to \varphi_1 + k^2 E_2\frac{\partial\varphi_1}{\partial E}, \tag{3.33b}$$

where the φ's and their derivatives on the right hand side are all evaluated at $E = E_0$. Henceforth it will be convenient to introduce the notation

$$\left\{\begin{array}{l} \varphi_0' = +\dfrac{\partial\varphi_0}{\partial\varepsilon} = -\dfrac{\partial\varphi_0}{\partial E} \\ \varphi_0'' = \dfrac{\partial^2\varphi_0}{\partial\varepsilon^2}, \qquad \varphi_l' = \dfrac{\partial\varphi}{\partial\varepsilon} = -\dfrac{\partial\varphi}{\partial E} . \end{array}\right. \qquad \varepsilon = -E \tag{3.34}$$

In order to get the energy values we expand the determinant of (3.31) in

powers of k, using (3.33) and the fact that $q_0(E_0) = 0$. From the coefficients of k^2 and k^4 we obtain, respectively:

(3.35) From coeff. of k^2: $$E_2 - \frac{1}{3} r_s^2 \frac{\varphi_1}{\varphi_0'},$$

(3.36) From coeff. of k^4: $$E_4 = \frac{2}{15} r_s^4 \frac{\varphi_1}{\varphi_0'} \left\{ 1 - \frac{2}{3}\frac{\varphi_1}{\varphi_2} - \frac{5}{6}\frac{\varphi_1'}{\varphi_0'} + \frac{5}{12}\frac{\varphi_0'' \varphi_1}{\varphi_0' \varphi_0'} \right\},$$

where again the reader is reminded that the φ's and their energy derivatives are to be evaluated at $E = E_0$. It is of interest to apply the empty lattice test to (3.35) and (3.36). To do this we make use of the following power series expansions of the spherical Bessel functions:

(3.37)
$$\begin{cases} j_0(x) = 1 - \frac{1}{6} x^2 + \frac{1}{120} x^4, \\ j_1(x) = \frac{1}{3} x - \frac{1}{30} x^3, \\ j_2(x) = \frac{1}{15} x^2 - \frac{1}{210} x^4. \end{cases}$$

If set $x = k r_s$, we find, using the definitions (3.34) and (3.32),

(3.38)
$$\begin{cases} \varphi_0' = \frac{1}{3} r_s^2 & \varphi_0'' = -\frac{2}{45} r_s^4 \\ \varphi_1 = 1 & \varphi_1' = \frac{1}{5} r_s^2 \\ \varphi_2 = 2 & \varphi_l = l. \end{cases}$$

From (3.38) it follows immediately that $E_2 = 1$ and $E_4 = 0$, which is, of course, the correct answer for the empty lattice.

TABLE XII. – *Sodium.*

	Empty Lattice	Metal	
φ_0	0	0	
φ_1	1.0	1.0046	
φ_0'	5.333	5.242	
$-\varphi_0''$	11.378	9.694	$E_0 \sim = -.6011$
φ_1'	3.200	3.300	
φ_2	2.000	2.034	
E_2	1.000	1.022	
E_4	0.00	−.0096	

It is also of interest to see how close the values for real metals come to the ideal free electron values of (3.38). Results for Na at $r_s = 4.0$ are summarized in Table XII. It is seen that all the relevant quantities are remarkably close to the empty lattice values, thus supporting the belief that Na is free electron like. It is also well known that Li is rather far from free electron like in its behavior, and this is confirmed by Table XIII. However, it is noteworthy that the large deviation from free electron behavior is associated with the E_2 term in the energy, and the coefficient E_4 is remarkably small. The smallness

TABLE XIII. – *Lithium.*

	Empty Lattice	Metal	
φ_0	0	0	
φ_1	1.0	0.6761	
φ_0'	3.413	3.159	
$-\varphi_0''$	4.638	3.381	$E_0 = -.6865$
φ_1'	2.048	2.582	
φ_2	2.0	1.930	
E_2	1.0	0.7305	
E_4	0.00	−.0303	

of E_4 is not an accident which happens to occur for the particular lattice constant chosen, since calculations for a wide range of lattice constants indicate that E_4 remains small over a wide range. For example, at $r_s = 2.4$, E_2 has fallen to 0.394, but E_4 is only $+0.009$.

The expression (3.35) was first derived by BARDEEN, but was given in the somewhat different form

$$E_2 = \gamma\varphi_1\,, \qquad \gamma = \frac{\frac{1}{3}r_s^3 f_0^2(r_s)}{\int_0^{r_s} f_0^2(r) r^2\,\mathrm{d}r}\,. \tag{3.39}$$

It is to be noted that γ involves a knowledge of the radial wave function over the entire volume of the cell, whereas (3.35) requires a knowledge only of surface values of the wave function. This apparent paradox is resolved with the aid of a theorem which was apparently first proved by BOWERS [18]. We consider the Schrödinger equation and differentiate it with respect to the energy parameter E, obtaining thus the pair of equations:

$$\begin{cases} \nabla^2\psi + (E - V)\psi = 0\,, \\ \nabla^2\dfrac{\partial\psi}{\partial E} + (E - V)\dfrac{\partial\psi}{\partial E} = -\psi\,. \end{cases} \tag{3.40}$$

We will assume ψ real, multiply the first equation by $\partial\psi/\partial E$, the second by ψ, subtract and integrate over the spherical cell, employing Green's theorem to transform the Laplacian term into a surface integral. In this way we obtain:

$$\int\left[\frac{\partial\psi}{\partial E}\frac{\partial\psi}{\partial r_s}-\psi\frac{\partial^2\psi}{\partial E\,\partial r_s}\right]\mathrm{d}S=\int\psi^2\,\mathrm{d}V\,, \tag{3.41}$$

where the $\mathrm{d}S$ represents surface integration, and $\mathrm{d}V$ volume integration. It is to be noted that equation (3.41) is valid only if both ψ and $\partial\psi/\partial E$ are well behaved at the origin. Thus it is tacitly assumed that ψ satisfies a boundary condition at the origin continuously as we vary E. This is important because it implies that information about the behavior of the wave function inside the cell is contained implicitly in the quantity $\partial\psi/\partial E$ and we are thus not getting something for nothing by expressing E_2 purely in terms of surface values of the wave function. However, as we shall see later, it is still not necessary to know ψ explicitly inside the cell in order to evaluate the left hand side of (3.41).

If we assume that ψ has the form

$$\psi=f_l(E,r)P_l(\cos\theta) \tag{3.42}$$

and carry out the angular integration, we obtain

$$r_s^2\left(\frac{\partial f_l}{\partial E}\frac{\partial f_l}{\partial r_s}-f_l\frac{\partial^2 f_l}{\partial E\,\partial r_s}\right)=\int_0^{r_s}f_l^2(r)r^2\,\mathrm{d}r\,, \tag{3.43}$$

we note that the left hand side may be written as

$$r_s^2f_l^2(r_s)\frac{\partial}{\partial\varepsilon}\frac{f_l'(r_s)}{f(r_s)}=r_sf_l^2(r_s)\varphi_l'\,, \tag{3.44}$$

using the definitions (3.34). From (3.44) it follows immediately that

$$\gamma=\tfrac{1}{3}r_s^2/\varphi_0' \tag{3.45}$$

and the identity of (3.35) and (3.39) follows immediately.

The expression (3.36) for E_4 is similarly not identical in form with that obtained by SILVERMAN, but SILVERMAN's result may be transformed into it with the aid of the theorem (3.44).

7. – Wave functions away from $k=0$ in the sphere approximation.

The results of Tables XI, XII and XIII suggest that near the cell boundaries the radial wave functions behave remarkably like spherical Bessel functions of argument $\sqrt{E-E_0}\,r_s$, i.e. the spherical harmonic components in

the expansion of a plane wave having a kinetic energy equal to the energy above the bottom of the valence band $E - E_0$. This suggests that if we express the Bloch functions in the form (3.13) the function $u(\boldsymbol{k}, r)$ ought to be very nearly constant near the cell boundary and to be quite insensitive to $\boldsymbol{k}$, for if we had an empty lattice $u(\boldsymbol{k}, \boldsymbol{r})$ would be exactly a constant independent of $\boldsymbol{k}$. If the Bloch function were a single orthogonalized plane wave, $u(\boldsymbol{k}, \boldsymbol{r})$ would still be a constant and very nearly independent of $\boldsymbol{k}$ near the boundary as long as the core was sufficiently small compared with the size of the cell. In fact $u(\boldsymbol{k}, \boldsymbol{r})$ would be nearly independent of $\boldsymbol{k}$ as long as $\boldsymbol{k}$ is small enough so that kr_c is small compared with l, where r_c is the core radius. This follows because when such a condition is satisfied the phase of the plane wave does not change appreciably across the core, and hence the non-orthogonality coefficient does not depend on $\boldsymbol{k}$. All of these considerations suggest that if $u_n(\boldsymbol{k}, \boldsymbol{r})$ instead of $b_n(k, r)$ is expanded in a power series in $\boldsymbol{k}$, this series ought to converge extremely rapidly, and it should be possible to represent the wave function quite accurately with only a few terms even for rather large values of $\boldsymbol{k}$. As we shall see this is in fact the case.

In order to find this expansion, we will first make some assumptions about the formal expansion (3.21). We will assume that it can be written in the form

$$b(k, r) = f_0 + ikr_s C_{11} f_1 P_1 + (ikr_s)^2 [C_{02} f_0 + C_{22} f_2 P_2] + \tag{3.46}$$

$$+ (ikr_s)^3 [C_{13} f_1 P_1 + C_{33} f_3 P_3] + (ikr_s)^4 [C_{04} f_0 + C_{24} f_2 P_2 + C_{44} f_4 P_4] +$$

$$+ \text{terms in } (ikr_s)^6, \quad \text{etc.}$$

In this expansion the f's are evaluated at the actual energy E, the P's stand for Legendre polynomials, and the C's are coefficients to be determined by the boundary conditions. Before we can complete the expansion in powers of k, we must do three things. First we must multiply (3.46) by $\exp[-ikr\cos\theta]$ in order to convert it to $u(k, r)$. Second in order to get u in the form of a power series, we must expand the exponential factor in a power series, multiply by the series (3.46), and collect terms of the same degree. Third we must carry out a Taylor's expansion of the f's about E_0 as in (3.33) and collect terms of the proper degree in k. If we further assume that all the f's are uniformly for all energies equal to 1 at the cell surface, we may express the value of u at the surface of the cell entirely in terms of the φ's. Carrying out these procedures we obtain the formal expansion:

$$u(\boldsymbol{k}, \boldsymbol{r}) = v_{00} + ikr_s v_{11} P_1 + (ikr_s)^2 [v_{02} + v_{22} P_2] + \tag{3.47}$$

$$+ (ikr_s)^3 [v_{13} P_1 + v_{33} P_3] + (ikr_s)^4 [v_{04} + v_{24} P_2 + v_{44} P_4],$$

where the v's are certain functions of r for the energy E_0. Using the procedure outlined above we find the following expressions for the v's:

$$
(3.48)\qquad \begin{cases} v_{00} = f_0\,, \qquad v_{22} = C_{22} f_2 + \dfrac{1}{3} f_0 \dfrac{r^2}{r_s^2} - \dfrac{2}{3} C_{11} f_1 \dfrac{r}{r_s}\,, \\ v_{11} = C_{11} f_1 - f_0 \dfrac{r}{r_s}\,, \\ v_{02} = C_{02} f_0 + \dfrac{1}{3} \dfrac{\varphi_1}{\varphi_0'} \dfrac{\partial f_0}{\partial \varepsilon} + \dfrac{1}{6} f_0 \dfrac{r^2}{r_s^2} - \dfrac{1}{3} C_{11} f_1 \dfrac{r}{r_s}\,, \end{cases}
$$

where we have retained only terms up to order 2 in k.

We now use the boundary conditions in the form (3.14), which for the spherical case take the simple forms:

$$
(3.49)\qquad \begin{cases} v_{00}' = v_{02}' = v_{22}' = 0\,, \\ v_{11} = 0\,. \end{cases}
$$

The first condition is authomatically satisfied if we take $E = E_0$. The condition on v_{11} gives $C_{11} = 1$. The condition on v_{02} is automatically satisfied, since we have used (3.35). The condition on v_{22} gives

$$
(3.50)\qquad C_{22}\varphi_2 - \frac{2}{3}\varphi_1 = 0\,, \qquad C_{22} = \frac{2}{3}\frac{\varphi_1}{\varphi_2}\,.
$$

Thus we have finally:

$$
(3.51)\qquad \begin{cases} v_{00} = f_0(r)\,, \qquad v_{11} = f_1(r) - \dfrac{r}{r_s} f_0(r)\,, \\ v_{02}(r) = C_{02} f_0(r) + \dfrac{1}{3} \dfrac{\varphi_1}{\varphi_0'} \dfrac{\partial f_0(r)}{\partial \varepsilon} + \dfrac{1}{6} \dfrac{r^2}{r_s^2} f_0(r) - \dfrac{1}{3} \dfrac{r}{r_s} f_1(r)\,, \\ v_{22}(r) = \dfrac{2}{3} \dfrac{\varphi_1}{\varphi_2} f_2(r) + \dfrac{1}{3} f_0(r) \dfrac{r^2}{r_s^2} - \dfrac{2}{3} \dfrac{r}{r_s} f_1(r)\,. \end{cases}
$$

We also note the values of these functions at the boundary:

$$
(3.52)\qquad \begin{cases} v_{00} = 1\,, \\ v_{11} = 0\,, \\ v_{02} = C_{02} - \dfrac{1}{6}\,, \\ v_{22} = \dfrac{2}{3} \dfrac{\varphi_1}{\varphi_2} - \dfrac{1}{3}\,. \end{cases}
$$

We observe that C_{02} is not determined by the boundary conditions. It is in fact arbitrary until we normalize the wave function in the cell. For the normalization it is sufficient to use (3.46) as long as we expand in powers of k. This follows because the normalization of u and b is the same, and if we expand each of their normalization integrals in powers of k, the coefficients of equal powers must be identical. To order k^2 the normalization integral becomes:

$$[1 - 2C_{02}k^2r_s^2]\int_0^{r_s} f_0^2(r)r^2\,dr + k^2r_s^2\frac{1}{3}\int_0^{r_s} f_1^2(r)r^2\,dr - 2k^2E_2\int_0^{r_s} f_0(r)\frac{\partial}{\partial\varepsilon}f_0(r)r^2\,dr\,. \tag{3.53}$$

We have from (3.43) and (3.44) the theorem

$$\int_0^s f_l^2(r)r^2\,dr = r_s\varphi_l'\,. \tag{3.54}$$

If we differentiate both sides with respect to energy we obtain

$$2\int_0^{r_s} f_l(r)\frac{\partial f_l(r)}{\partial\varepsilon} = r_s\varphi_l''\,. \tag{3.55}$$

The condition that the normalization be independent of k to order k^2 then leads to:

$$C_{02}(r_s) = \frac{1}{6}\frac{\varphi_1'}{\varphi_0'} - \frac{1}{6}\frac{\varphi_1}{\varphi_0'}\frac{\varphi_0''}{\varphi_0'}\,, \tag{3.56}$$

whence from (3.52) we obtain:

$$v_{02}(r_s) = -\frac{1}{6}\left\{1 - \frac{\varphi_1'}{\varphi_0'} + \frac{\varphi_1}{\varphi_0'}\frac{\varphi_0''}{\varphi_0'}\right\}. \tag{3.57}$$

Again we notice, using (3.38), that for the free electron case v_{02} vanishes at the boundary. Furthermore, the functions in (3.51) vanish identically at all points in the cell for an empty lattice. If we remember that the f's in (3.51) are defined so that they equal 1 on the cell boundary for all energies, and that they are evaluated at the energy for the ground state, which in this case is zero, it follows that:

$$\begin{cases} f_0(r) &= 1 \\ f_1(r) &= \dfrac{r}{r_s}\,, \\ f_2(r) &= \dfrac{r^2}{r_s^2}\,, \\ \dfrac{\partial f_0(r)}{\partial\varepsilon} &= \dfrac{1}{6}(r^2 - r_s^2)\,. \end{cases} \tag{3.58}$$

These results substituted in (3.51) show that v_{00} is constant and equal to 1 while all the other v's vanish.

Table XIV shows the values of the v's for Na and Li, using the values for the q's given in Tables XII and XIII. It is also of interest to note how the

TABLE XIV.

Quantity	Na ($r_s = 4.0$)	Li ($r_s = 3.2$)
$v_{02}(r_s)$	$-.0026$	$+.008$
$v_{22}(r_s)$	$-.00406$	$-.100$

v's behave in the neighborhood of $r = r_s$, and to this end we show in eq. (3.59) an expression for the lowest non-vanishing derivatives of the v's in (3.51) evaluated for $r = r_s$. In deriving these expressions use has been made of the fact that the f's satisfy radial differential equations for the appropriate l-value,

$$r_s^2 v_0''(r_s) = r_s^2(V - E_0)\,, \tag{3.59a}$$

$$r_s v_{11}' = \varphi_1 - 1\,, \tag{3.59b}$$

$$\begin{cases} r_s^2 v_{22}''(r_s) = \left(\dfrac{2}{3}\dfrac{\varphi_1}{\varphi_2} - \dfrac{1}{3}\right) r_s^2\,(V - E_0) + \left(4\dfrac{\varphi_1}{\varphi_2} - \dfrac{2}{3} - \dfrac{4}{3}\varphi_1\right) = \\ \qquad = r_s^2(V - E_0)\, v_{22}(r_s) + \left(4\dfrac{\varphi_1}{\varphi_2} - \dfrac{2}{3} - \dfrac{4}{3}\varphi_1\right), \\ r_s^2 v_{02}''(r_s) = -\dfrac{1}{6} r_s^2(V - E_0)\left[1 - \dfrac{\varphi_1'}{\varphi_0'} + \dfrac{\varphi_2'}{\varphi_0'}\dfrac{\varphi_0''}{\varphi_0'}\right] + \dfrac{1}{3}(E_2 - 1) = \\ \qquad = r_s^2(V - E_0) + v_{02}(r_s) + \dfrac{1}{3}(E_2 - 1)\,. \end{cases} \tag{3.59c}$$

Numerical values for these expressions are shown for Na and Li in Table XV, where it is assumed that $V(r) = -2/r$ near the cell boundary.

TABLE XV.

Quantity	Na	Li
$r_s v_{11}'(r_s)$	$+.0046$	$-.3239$
$r_s^2 v_{02}''(r_s)$	$-.0027$	$-.103$
$r_s^2 v_{22}''(r_s)$	$-.0371$	$-.232$

It is seen from Tables XIV and XV that the convergence of the series is very rapid for Na but not as good for Li. In considering the rate of con-

vergence it is worth remembering that at the Fermi surface we have

$$k_0 r_s = \left(\frac{9\pi}{4}\right)^{\frac{1}{3}} \simeq 1.92 \,. \tag{3.60}$$

Thus at the Fermi surface of Na $v_{22}(r_s)$ is about 1.5% of $v_{00}(r_s)$, which is unity by definition. On the other hand, for Li it is nearly 40%, so that convergence is really very poor.

In order to examine the convergence still further we give below the expression for $v_{44}(r_s)$,

$$v_{44}(r_s) = \frac{1}{21} - \frac{4}{35}\frac{\varphi_1}{\varphi_2} - \frac{8}{105}\frac{\varphi_1}{\varphi_4} - \frac{8}{105}\frac{\varphi_3}{\varphi_4} + \frac{8}{35}\frac{\varphi_1}{\varphi_2}\frac{\varphi_3}{\varphi_4} \,. \tag{3.61}$$

The numerical values are $2.6 \cdot 10^{-5}$ for Na and $-.0029$ for Li. At the Fermi surface the Li value is about 4% the value of $v_{00}(r_s)$, so the convergence for Li apparently becomes good after the first term, but not nearly as good as for Na. The reason for the improvement in convergence with the higher terms can be seen if we remember that the higher φ's assume values close to their free electron values. If we assume that only φ_1 deviates appreciably from the free electron value, then (3.61) becomes approximately:

$$v_{44}(r_s) \approx \frac{\varphi_1 - 1}{103} + \frac{\varphi_1}{35}\left(\frac{2}{\varphi_2} - 1\right) . \tag{3.62}$$

It is seen that the deviations of both φ_1 and φ_2 from their free electron values are divided by rather large numerical factors in the higher terms in the expansion, so that even though these deviations may be large they apparently do not make a large contribution after the leading term of the expansion to which they contribute.

If we assume

$$\varphi_l = l(1 + \sigma_l) \tag{3.63}$$

where σ_l is small then we may write approximately:

$$v_{44}(r) = \frac{1}{35}\left(\frac{\sigma_1 - \sigma_4}{3} + \sigma_3 - \sigma_2 ,\right) \tag{3.64}$$

which again shows the smallness of the term. Eventually the φ's are bound to become free-electron like as the centrifugal potential prevents the electron from penetrating into the region where the electrostatic potential is appreciable. This point will be discussed at greater length when we come to deal with the calculation of the radial wave functions in Chapter V.

REFERENCES

[1] J. R. Reitz: *Solid State Physics*, vol. I (New York, 1956), pp. 1-95, cf. especially p. 20; original paper, F. Bloch: *Zeits. Phys.*, **52**, 555 (1928); F. Seitz: *Modern Theory of Solids* (New York, 1940), chap. 8.

[2] For example cfr. J. R. Reitz: *Solid State Physics*, vol. I (New York, 1956), pp. 1-95, cfr. especially p. 20; the basic original papers are the following: F. Seitz: *Z. F. Kristallographie*, (A)**88**, 433 (1943); **90**, 289 (1935); **91**, 336 (1935); **94**, 100 (1936); *Ann. of Math.*, **37**, 17 (1936); C. Herring: *Phys. Rev.*, **61**, 365 (1937); *Journ. Franklin Institute*, **233**, 525 (1942); L. P. Brouckaert, R. Smoluchowski and E. P. Wigner: *Phys. Rev.*, **50**, 58 (1936). For a review cfr. also D. G. Bell: *Rev. Mod. Phys.*, **26**, 311 (1954).

[3] R. J. Elliott: *Phys. Rev.*, **96**, 266 (1954); G. Dresselhaus: *Phys. Rev.*, **100**, 580 (1955); R. H. Parmenter: *Phys. Rev.*, **100**, 573 (1955).

[4] J. C. Slater: *Phys. Rev.*, **45**, 794 (1934); *Rev. Mod. Phys.*, **6**, 209 (1934).

[5] W. Shockley: *Phys. Rev.*, **52**, 866 (1937).

[6] F. C. Von der Lage and H. A. Bethe: *Phys. Rev.*, **71**, 612 (1947).

[7] D. J. Howarth and H. Jones: *Proc. Phys. Soc.*, A **65**, 355 (1952).

[8] C. Herring: *Phys. Rev.*, **57**, 1169 (1940); C. Herring and A. G. Hill: *Phys. Rev.*, **58**, 132 (1940).

[9] F. Herman: *Phys. Rev.*, **93**, 1214 (1954).

[10] J. Callaway: *Phys. Rev.*, **97**, 933 (1955).

[11] R. H. Parmenter: *Phys. Rev.*, **86**, 552 (1952).

[12] W. Kohn: *Phys. Rev.*, **87**, 472 (1952).

[13] L. Pincherle: *Physica*, **20**, 965 (1954).

[14] E. P. Wigner and F. Seitz: *Phys. Rev.*, **43**, 804 (1933).

[15] J. Bardeen: *Journ. Chem. Phys.*, **6**, 367 (1938).

[16] R. A. Silverman: *Phys. Rev.*, **85**, 227 (1952).

[17] W. Kohn: *Phys. Rev.*, **96**, 590 (1954).

[18] W. A. Bowers: *Thesis* (Cornell University, 1943).

[19] For a good elementary discussion of the properties of crystal lattices see C. F. Kittel: *Introduction to Solid State Physics*, 2nd ed. (New York, 1956), chap. I.

[20] F. S. Ham: *Thesis* (Harvard University, 1954).

IV.

Solution of the one electron problem: energy bands, continued.

1. – The boundary correction integral.

Let us assume we have an exact solution of the Schrödinger equation which satisfies the boundary conditions only approximately. Howarth and Jones [1] have given a procedure by which the energy may be corrected. We shall not give the Howarth and Jones proof but rather an equivalent one which is

more appropriate to the (3.14) form of the B.C. We note that if the Bloch function is written in the form (3.13) the function $u(\boldsymbol{k}, \boldsymbol{r})$ satisfies the differential equation [2]

$$\nabla^2 u + (E - V)u + 2i\boldsymbol{k}\cdot\nabla u - k^2 u = 0 \,. \tag{4.1}$$

Let us now consider two solutions of (4.1), one, labelled u^*, which satisfies the B.C. exactly, and is the complex conjugate of the solution shown in (4.1) and satisfies the complex conjugate equation; the other, labelled v, which satisfies (4.1) but fulfills only approximate B.C., e.g. spherical B.C. We will let the energy for the exact solution be $E+\Delta E$, and that for the approximate solutions E. We then have:

$$\begin{cases} \nabla^2 u^* + (E + \Delta E - V(r) - k^2)u^* - 2i\boldsymbol{k}\cdot\nabla u^* = 0 \,, \\ \\ \nabla^2 v + (E - V(r) - k^2)v + 2i\boldsymbol{k}\cdot\nabla v = 0 \,. \end{cases} \tag{4.2}$$

We multiply the first equation by v and second by u^*, subtract, and integrate over the atomic polyhedron, using Green's theorem to convert the Laplacian terms into a surface integral. We thus have:

$$\Delta E \int u^* v \, \mathrm{d}\tau = \int \left[u^* \frac{\partial v}{\partial n} - v \frac{\partial u^*}{\partial n} \right] \mathrm{d}S + 2i\boldsymbol{k}\cdot \int \hat{n} u^* v \, \mathrm{d}S. \tag{4.3}$$

On the left-hand side we will not make much error by taking $u^* = v^*$ and replacing the polyhedron by a sphere of equal volume. On the right-hand side we proceed as follows. Let primed quantities denote values of the functions at conjugate points on the atomic polyhedron, i.e., at the points which correspond in the B.C. In this notation the B.C. (3.14) may be written:

$$u' = u \,, \qquad \frac{\partial u'}{\partial n'} = -\frac{\partial u}{\partial n} \,, \tag{4.4.}$$

where the normals n and n' are oppositely directed. We carry out the surface integrals by pairing off conjugate points on the polyhedron. In this way the right side of (4.3) may be written

$$\frac{1}{2}\int \left[u^{*\prime} \frac{\partial v'}{\partial n'} + u^* \frac{\partial v}{\partial n} - v' \frac{\partial u^{*\prime}}{\partial n'} - v \frac{\partial u^*}{\partial n} \right] \mathrm{d}S + i\boldsymbol{k}\cdot \int [\hat{n}' u^{*\prime} v' + \hat{n} u^* v] \, \mathrm{d}S \,. \tag{4.5}$$

Using the B.C. (4.4) eq. (4.5) may be rewritten:

$$\frac{1}{2}\int \left[u^* \left(\frac{\partial v'}{\partial n'} + \frac{\partial v}{\partial n} \right) + \frac{\partial u^*}{\partial n} (v' - v) \right] \mathrm{d}S - i\boldsymbol{k}\cdot \int \hat{n} u^* (v' - v) \, \mathrm{d}S \,. \tag{4.6}$$

We now suppose that we may write $u = v - \omega$, where ω is a quantity which may be considered small of the first order. We note that because of (4.4) we have

$$\frac{\partial v'}{\partial n'} + \frac{\partial v}{\partial n} = \frac{\partial \omega'}{\partial n'} + \frac{\partial \omega}{\partial n}, \qquad v' - v = \omega' - \omega. \tag{4.7}$$

It follows that (4.6) may be written

$$\frac{1}{2}\int \left[v^* \left(\frac{\partial v'}{\partial n'} + \frac{\partial v}{\partial n} \right) + \frac{\partial v^*}{\partial n} (v' - v) \right] \mathrm{d}S - i\boldsymbol{k} \cdot \int \hat{n} v^* (v' - v)\, \mathrm{d}S , \tag{4.8}$$

plus terms of order ω^2, which we will assume can be neglected compared with those of the first order in ω.

From the expansion (3.47) it follows that the real part of v contains only even harmonics and the imaginary part of v contains only odd harmonics. If the polyhedral cell has a center of inversion, as it does in a Bravais lattice, then it is possible to pair off points on the surface of the cell which are related by inversion in the center of the cell. By doing this and using the even or odd property of the harmonics with respect to inversion it is readily shown that all the terms in the integrand (4.8) which do not contain both a primed and an unprimed term cancel, and we are left with:

$$\Delta E \int |v|^2 \,\mathrm{d}\tau = \frac{1}{2} \int \left[v' \frac{\partial v^*}{\partial n} - v^* \frac{\partial v'}{\partial n} \right] \mathrm{d}S - i\boldsymbol{k} \cdot \int n v^* v' \,\mathrm{d}S . \tag{4.9}$$

This equation is the required expression for the boundary perturbation.

2. – Application to the lowest valence state.

If $\boldsymbol{k} = 0$, we may take v as a function of radius only. Equation (4.9) then becomes

$$4\pi \Delta E \int_0^{r_s} f_0^2(r) r^2 \,\mathrm{d}r = \frac{1}{2} \int f_0(r) \frac{\mathrm{d}f_0}{\mathrm{d}r}(r)\, \hat{n} \cdot (\hat{r} - \hat{r}')\, \mathrm{d}S \tag{4.10}$$

where r, r' are unit vectors in directions to conjugate points on the cell face, the geometry of which is represented in Fig. 2. From the figure we see that

$$\hat{n} \cdot (\hat{r} - \hat{r}') = 2\hat{n} \cdot \hat{r} . \tag{4.11}$$

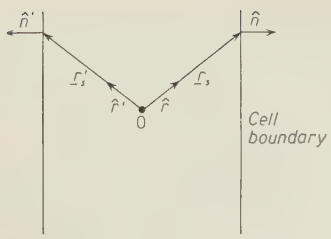


Fig. 2. – Definition of boundary vectors for cellular boundary conditions.

To evaluate the integral we make a Taylor expansion of the integrand about $r = r_s$, using the following results:

$$(4\cdot12)\qquad v_{00}(r) = 1 + \frac{1}{2}\left(\frac{r}{r_s} - 1\right)^2 r_s^2[V(r_s) - E_0] =$$

$$= \left(\frac{r}{r_s} - 1\right) r_s[V(r_s) - E_0] + \frac{1}{2}\left(\frac{r}{r_s} - 1\right)^2 r_s^2 f_0'''(r_s)\,,$$

where from the differential equation

$$(4.13)\qquad r_s^2 f_0'''(r_s) = -2r_s[V(r_s) - E_0] + r_s^2\frac{\mathrm{d}V}{\mathrm{d}r_s} = 6 - 2r_s\varepsilon_0\,,$$

the last result being for a Coulomb potential. If we substitute this result in (4.10) we have:

$$(4.14)\qquad 4\pi r_s\varphi_0'\Delta E = r_s[V(r_s) - E_0]\int\left(\frac{r}{r_s} - 1\right)(\hat{n}\cdot\hat{r})\,\mathrm{d}S +$$

$$+ \left\{-r_s[V(r_s) - E_0] + \frac{1}{2}r_s^2\frac{\mathrm{d}V}{\mathrm{d}r_s}\right\}\int\left(\frac{r}{r_s} - 1\right)^2(\hat{n}\cdot\hat{r})\,\mathrm{d}S\,.$$

The surface integrals are purely geometrical quantities, depending only on the geometry of the atomic polyhedron and may be evaluated once and for all for a given crystal structure. Both integrals are of the same order of magnitude because in the first integral $(r/r_s - 1)$ may have either sign and the two signs tend to cancel against each other. In fact, it can be shown that if r_s is chosen so that the volume of the atomic polyhedron and that of the sphere are equal, the first integral vanishes to a very high approximation.

For the case of a Coulomb potential near the cell boundary (4.14) may be put in the form

$$(4.15)\qquad \Delta E = \frac{r_s}{\varphi_0'}(3 - \varepsilon_0 r_s)\frac{S}{4\pi r_s^2}\frac{1}{S}\int\left(\frac{r}{r_s} - 1\right)^2(\hat{n}\cdot\hat{r})\,\mathrm{d}S\,,$$

where S is the total area of the polyhedron. The last two factors are now dimensionless and purely geometrical. We have

$$(4.16)\qquad \frac{S}{4\pi r_s^2} = 1.098\,,$$

$$(4.17)\qquad \frac{1}{S}\int\left(\frac{r}{r_s} - 1\right)^2(\hat{n}\cdot\hat{r})\,\mathrm{d}S = 00384\,.$$

For Na this gives a correction of $+.0020$ Ry in the ground state energy, or about 0.6 kcal/mol decrease in cohesive energy. If we had used the self-consistent potential, taking into account the field of an electron in the cell, we would have had $dV/dr_s = 0$ and $-r_s[V(r_s) - E_0] = -.0039$. This would have given over 1000 times smaller results for ΔE. These results thus justify the particular selection made by WIGNER and SEITZ for the size of the spherical cell. Furthermore, it is seen that what small correction there is is extremely sensitive to the precise choice of potential in the immediate neighborhood of the cell boundary. It is also interesting that the correction depends only on the properties of the potential and not directly on the wave function.

Actually even the result with the Coulomb potential cannot be regarded as reliable because we have neglected a correction due to the fact that in this case the true potential is in fact not spherically symmetric outside the inscribed sphere. It is easily shown that the true potential is everywhere lower than or equal to the assumed Coulomb potential, so that the effect of this correction is in a direction to cancel the correction (4.15). For this reason we believe that in effect the correction to the ground state energy is entirely negligible. Another test example is obtained by taking the potential constant outside the inscribed sphere of the atomic polyhedron. For this case, we can easily estimate $\Delta E = -.00054$ Ry.

3. – Application to correction of E_2:

When $\boldsymbol{k}$ differs from 0 we must collect all the terms in (4.9) which are of the same order in $\boldsymbol{k}$. For the terms of degree 2 in k, the following pairs of expansion functions will make contributions to the surface integrals in (4.9):

$$\text{(4.18)} \qquad \begin{cases} v_{00} \text{ and } v_{02}\,, \\ v_{00} \text{ and } v_{22}\,, \\ v_{11} \text{ and } v_{11}\,, \\ v_{00} \text{ and } v_{11}\,. \end{cases}$$

The last pair contributes only via the second integral on the right of (4.9). In evaluating the integrand we make use of the following general formula for the normal derivatives:

$$\text{(4.19)} \qquad \hat{n}\cdot\nabla v_{ln}(r)\,P_l(\hat{k}\cdot\hat{r}) = (\hat{n}\cdot\hat{r})\,P_l(\hat{k}\cdot\hat{r})\,\frac{\partial v_{ln}}{\partial r} + \frac{v_{ln}(r)}{r}\,P_l'(\hat{k}\cdot\hat{r})[(\hat{k}\cdot\hat{n}) - (\hat{k}\cdot\hat{r})(\hat{n}\cdot\hat{r})]\,,$$

where $P_l'(\hat{k}\cdot\hat{r})$ represents the derivative of the Legendre polynomial with respect to its argument. When we substitute (4.19) into the integrand of (4.9)

we obtain rather complicated expressions which can be greatly simplified by considerations of symmetry. Here we shall confine our attention for the moment to the b.c.c. structure. Let T_i be a transformation appropriate to cubic symmetry, and let us denote by $T_i f(x, y, z)$ the function obtained by applying the transformation to the argument, i.e.

$$T_i f(\boldsymbol{r}) = f(T_i \boldsymbol{r}) , \tag{4.20}$$

where $T_i \boldsymbol{r}$ is the transformed co-ordinate vector. We then have for a surface integral on a figure of cubic symmetry

$$\int f(\boldsymbol{r})\, \mathrm{d}S = \frac{1}{N} \int \sum_i T_i f(\boldsymbol{r})\, \mathrm{d}S , \tag{4.21}$$

where the summation is taken over all the operations which leave the cell unchanged, in our case the 48 operations of the cubic group.

Using these results we obtain finally:

$$\Delta E \int |v|^2 \,\mathrm{d}\tau = k^2 r_s^2 \Big\{ \int [v_{00} v'_{02} + v_{02} v'_{00}](\hat{n}\cdot\hat{r})\,\mathrm{d}S + \tag{4.22}$$

$$+ \frac{1}{3} \int r_{11} r'_{11} (\hat{n}\cdot\hat{r})[1 - 2(\hat{n}\cdot\hat{r})^2]\,\mathrm{d}S - \frac{2}{3} \int \frac{v_{11}^2}{r} (\hat{n}\cdot\hat{r})[1 - (\hat{n}\cdot\hat{r})^2]\,\mathrm{d}S - \frac{2}{3} \int \frac{v_{00} v_{11}}{r_s} (\hat{n}\cdot\hat{r})\,\mathrm{d}S \Big\} .$$

In this expression the v's are to be evaluated at the radius appropriate to the polyhedral surface, and it is to be observed that the primes in (4.22) denote differentiation with respect to r, *not* conjugate points on the surface. The quantity $(\hat{n}\cdot\hat{r})$ is the cosine of the angle between the radius vector to a point on the polyhedral surface and the normal to the surface. For symmetrical surfaces it is always very close to unity for all points on the surface. The normalization integral on the left hand side should contain terms of order k^2, but by (3.52) and (3.56) we have determined C_{02} and hence v_{02} so that the normalization is independent of k to order k^2. Thus the left hand side may be replaced by its value at $k = 0$ in the present approximation.

In order to take advantage of the fact that $(\hat{r}\cdot\hat{n})$ is nearly unity, we rewrite (4.22) in the form:

$$\Delta E_2 = \frac{r_s}{\varphi'_0} \Big\{ \int \Big[v_{00} v'_{02} + v_{02} v'_{02} - \frac{1}{3} v_{11} v'_{11} - \frac{2}{3} \frac{v_{00} v_{11}}{r_s} \Big] (\hat{n}\cdot\hat{r})\,\mathrm{d}S + \tag{4.23}$$

$$+ \int \Big[\frac{2}{3} v_{11} v'_{11} - \frac{2}{3} \frac{v_{11}^2}{r} \Big] (\hat{n}\cdot\hat{r})[1 - (\hat{n}\cdot\hat{r})^2]\,\mathrm{d}S .$$

The expression (4.23) can be simplified further by using Taylor expansions of the v's about $r = r_s$ as we did in the case of ΔE_0. If we carry this out, taking

into account that for the proper choice of sphere radius

$$\int\left(\frac{r}{r_s}-1\right)(\hat{n}\cdot\hat{r})\,\mathrm{d}S=0 \tag{4.24}$$

we obtain finally

$$\begin{aligned}\Delta E_2=\frac{r_s^2}{\varphi_0'}\frac{S}{4\pi r_s^2}\Big\{&\Big[\frac{1}{2}r_s^3v_{02}'''+\frac{1}{2}r_s^2v_{0-}v_{00}''+\frac{1}{2}r_s^3v_{11}'v_{11}''+\frac{1}{3}r_s^2v_{11}''-\frac{2}{3}r_s^2v_{11}'^2\Big]\cdot\\ &\cdot\frac{1}{S}\int\left(\frac{r}{r_s}-1\right)^2(\hat{n}\cdot\hat{r})\,\mathrm{d}S-\\ &-\Big[\frac{2}{3}r_s^2r_{11}'^2\Big]\frac{1}{S}\int\left(\frac{r}{r_s}-1\right)(\hat{n}\cdot\hat{r})^3\,\mathrm{d}S+\Big[\frac{2}{3}r_s^2r_{11}'^2-r_s^3r_{11}'v_{11}''\Big]\frac{1}{S}\int\left(\frac{r}{r_s}-1\right)^2(\hat{n}\cdot\hat{r})^3\,\mathrm{d}S\Big\}.\end{aligned} \tag{4.25}$$

The expressions in (4.25) can be simplified still further by utilizing the differential equations, and we obtain finally the result

$$\begin{aligned}\Delta E_2=\frac{r_s^2}{\varphi_0'}\frac{S}{4\pi r_s^2}\Big\{&\Big[\frac{v_{02}}{2}\{r_s^3V'-r_s^2(V-E_0)\}-(\varphi_1-1)-\frac{5}{3}(\varphi_1-1)^2+\\ &+\varphi_1(1-\gamma)\Big]\cdot\frac{1}{S}\int\left(\frac{r}{r_s}-1\right)^2(\hat{n}\cdot\hat{r})\,\mathrm{d}S-\frac{2}{3}(\varphi_1-1)^2\frac{1}{S}\cdot\\ &\cdot\int\left(\frac{r}{r_s}-1\right)(\hat{n}\cdot\hat{r})^3\,\mathrm{d}S+\frac{8}{3}(\varphi_1-1)^2\frac{1}{S}\int\left(\frac{r}{r_s}-1\right)^2(\hat{n}\cdot\hat{r})^3\,\mathrm{d}S\Big\}.\end{aligned} \tag{4.26}$$

The only term which depends explicitly upon the potential is of the same general form as in ΔE_0, except that it is multiplied by v_{02}, which is extremely small. Hence we may neglect this contribution entirely. The correction to E_0 vanishes as it should for the empty lattice, and for Na it is of order 10^{-5}, which is entirely negligible. Even for Li the term is only of order 10^{-3} and can be neglected, primarily because the surface integrals are so small.

The conclusion from the discussion of the last two sections is that the corrections to the E_0 and E_2 parts of the energy may be entirely neglected.

4. – The correction of E_4.

As might be expected the correction of E_4 involves many more types of terms than the correction of E_2, and it is also necessary to take into account functions up to the type v_{44} if we are to obtain a complete solution of the problem. Another complication is that there are in addition to the isotropic terms in k^4 some anisotropic terms which give rise to a warping of the spherical energy surfaces. The warping of the surfaces is, however, not of much interest

for cohesion unless it is very large, because when we average the energy over all directions in k space the anisotropic term averages out in the first approximation and makes no contribution to the total energy of the system. We will therefore confine our attention to the isotropic terms and shall calculate only a typical term to obtain the order of magnitude of the correction. As our illustration we shall choose the terms which arise from v_{22} associated with c_{22},

$$\Delta E_4 = \frac{r_s^4}{\varphi_0'}\frac{S}{4\pi r_s^2}\left\{\frac{1}{S}\int\frac{57}{40}\,n\left(\frac{39}{19}\,n^2-1\right)(1-n^2)\frac{r_s}{r}v_{22}^2\,\mathrm{d}S + \right. \tag{4.27}$$

$$\left. + \frac{1}{S}\int\frac{3}{40}\,n(9n^2-1)(1-n^2)v_{22}r_s v_{22}'\,\mathrm{d}S\right\},$$

where we have used n as an abbreviation for $(\hat{n}\cdot\hat{r})$. With the usual technique of Taylor's expansion of the wave functions we obtain:

$$\Delta E_4 = \frac{r_s^4}{\varphi_0'}\frac{S}{4\pi r_s^2}\left\{\left(\frac{2}{3}\frac{\varphi_1}{\varphi_2}-\frac{1}{3}\right)^2\frac{1}{S}\int\frac{57}{40}\,n(1-n^2)\left(\frac{39}{19}\,n^2-1\right)\mathrm{d}S + \right. \tag{4.28}$$

$$+\left[r_s^2(V-E_0)\left(\frac{2}{3}\frac{\varphi_1}{\varphi_2}-\frac{1}{3}\right)+2\left(2\frac{\varphi_1}{\varphi_2}-1\right)+\frac{4}{3}(1-\varphi_1)\right]\left(\frac{2}{3}\frac{\varphi_1}{\varphi_2}-\frac{1}{3}\right)\cdot$$

$$\left.\cdot\frac{1}{S}\int\left(\frac{r}{r_s}-1\right)\frac{3}{40}\,n(1-n^2)(9n^2-1)\,\mathrm{d}S\right\}.$$

The integrands in the angular integrals may be written:

$$\frac{57}{40}\,n(1-n^2)\left(\frac{39}{19}\,n^2-1\right)=\frac{3}{2}\,n(1-n^2)-\frac{117}{40}\,n(1-n^2)^2\,, \tag{4.29}$$

$$\frac{3}{40}\,n(1-n^2)(9n^2-1)=\frac{3}{5}\,n(1-n^2)-\frac{27}{40}\,n(1-n^2)^2\,. \tag{4.30}$$

Table XVI gives the values of the surface integrals. Table XVII shows values of the coefficients computed for the 5 alkali metals with neglect of $V-E_0$. In all cases the contribution of the second angular integral is negligible compared with the first, even though its coefficient is somewhat larger. Table XVIII shows the contribution to E_4 resulting from the numerical values shown in Tables XVI

TABLE XVI.

$\frac{1}{S}\int\left[\frac{3}{2}\,n(1-n^2)-\frac{117}{40}\,n(1-n^2)^2\right]\mathrm{d}S$	+ 0.1238
$\frac{1}{S}\int\left(\frac{r}{r_s}-1\right)\left[\frac{3}{5}\,n(1-n^2)-\frac{27}{40}\,n(1-n^2)^2\right]\mathrm{d}S$	− 0.000339

TABLE XVII.

	$\left(\frac{2}{3}\frac{\varphi_1}{\varphi_2} - \frac{1}{3}\right)^2$	$\left[2\left(2\frac{\varphi_1}{\varphi_2} - 1\right) + \frac{4}{3}(1 - \varphi_1)\right]\left(\frac{2}{3}\frac{\varphi_1}{\varphi_2} - \frac{1}{3}\right)$
Li ($r_s = 3.2$)	+ .00995	+ .0166
Na ($r_s = 4.0$)	$+1.66 \cdot 10^{-5}$	$+1.24 \cdot 10^{-4}$
K ($r_s = 4.8$)	+ .0021	+ .0106
Rb ($r_s = 5.2$)	+ .00576	+ .0230
Cs ($r_s = 5.4$)	+ .0210	+ .1040
($r_s = 5.8$)	+ .0127	+ .0544

and XVII. In every case it is interesting to note that the correction is in such a direction as to make E_4 smaller in magnitude, i.e. to make it approach the free electron value. Indeed for Li the correction is large enough to produce a change in sign. What we have tabulated, of course, is only one of several contributions to the correction in E_4, although probably the largest one. Thus the results of Table XVIII cannot be used quantitatively.

TABLE XVIII.

Element	r_s	$\frac{5}{7}\frac{E_4}{E_2} k_0^2$	$\frac{\Delta E_4}{E_4}$
Li	3.2	— .0106	— 1.36
Na	4.0	— .00155	— .014
K	4.8	— .032	— .0619
Rb	5.2	— .036	— .1410
Cs	5.4	— .086	— .214
	5.8	— .068	— .158

5. – Further improvements to the spherical approximation [3].

In Sect. **2**, **3** and **4** we have been discussing the boundary correction essentially as a first order perturbation, i.e. we have assumed the wave functions of the spherical approximation to be correct and then computed the first order perturbation of the energy with these wave functions using the boundary correction integral. This procedure is satisfactory provided that there are not two or more solutions in the spherical approximation that possess the same or nearly the same energy, and which at the same time may interact with each other as a result of the boundary perturbation.

We have seen that as long as we stick with spherical boundary conditions

it is meaningful to speak of Bloch functions which have values of k lying outside the first BZ. In other words we may still write functions in the form

$$\psi_n(\boldsymbol{r}) = \exp[i(\boldsymbol{k} + \boldsymbol{K}_n)\cdot\boldsymbol{r}]v(\boldsymbol{k} + \boldsymbol{K}_n, \boldsymbol{r}) , \tag{4.31}$$

where v satisfies the spherical boundary conditions but not necessarily the correct boundary conditions. Under these circumstances we cannot say that $v(\boldsymbol{k}, \boldsymbol{r})$ is periodic in k-space, satisfying (3.8), as we could if it were a true Bloch function.

We can construct Bloch functions out of the functions (4.31) by writing them in the form

$$b(\boldsymbol{k}, \boldsymbol{r}) = \exp[i\boldsymbol{k}\cdot\boldsymbol{r}]v(\boldsymbol{k}, \boldsymbol{r} - \boldsymbol{R}) . \tag{4.32}$$

In this equation $\boldsymbol{k}$ is an unrestricted wave vector, $\boldsymbol{r}$ is measured with respect to a fixed origin in the crystal, while the argument $\boldsymbol{r} - \boldsymbol{R}$ of $v(\boldsymbol{k}, \boldsymbol{r} - \boldsymbol{R})$ is always measured from the origin of the particular cell in which $\boldsymbol{r}$ lies. The function (4.32) now satisfies the Bloch condition (3.2) but is not continuous with continuous derivatives throughout the crystal.

We now consider two functions $b(\boldsymbol{k}, \boldsymbol{r})$ and $\varphi(\boldsymbol{k}_0, \boldsymbol{r})$, the first being defined by (4.32) and satisfying the Bloch condition but not the condition of continuity between cells, the second being the exact crystal function satisfying the Bloch condition for $\boldsymbol{k}_0$, and being continuous throughout the crystal. We may write

$$\nabla^2 b^*(\boldsymbol{k}, \boldsymbol{r}) - V(\boldsymbol{r})b^*(\boldsymbol{k}, \boldsymbol{r}) = -E_k b^*(\boldsymbol{k}, \boldsymbol{r}) , \tag{4.33a}$$

$$\nabla^2 \varphi(\boldsymbol{k}_0, \boldsymbol{r}) - V(\boldsymbol{r})\varphi(\boldsymbol{k}_0, \boldsymbol{r}) = -E\varphi(\boldsymbol{k}_0, \boldsymbol{r}) , \tag{4.33b}$$

where E_k is the energy value appropriate to the differential equation satisfied by $b(\boldsymbol{k}, \boldsymbol{r})$ inside any cell, and where E is the eigenvalue for the exact eigenfunction of the problem. We multiply a) by $\varphi_{\boldsymbol{k}_0}$ and b) by $b^*_{\boldsymbol{k}}$ subtract, and integrate over the whole volume of the crystal exclusive of an infinitesima[1] region which encloses all the cell boundaries, i.e. a region comprising the in[-] teriors of the cells only. The resulting volume and surface integrals may b[e] converted into sums of integrals over unit cell volumes and surfaces and finall[y] expressed in terms of integrals for a single unit cell, taking into account th[e] fact that both $\varphi_{\boldsymbol{k}_0}$ and $b^*_{\boldsymbol{k}}$ satisfy Bloch conditions of the form (3.2). In thi[s] way we obtain the condition:

$$\sum_{\boldsymbol{R}} \exp[i(\boldsymbol{k} - \boldsymbol{k}_0)\cdot\boldsymbol{R}] \int_{\text{cell}} \left\{ \varphi_{\boldsymbol{k}_0} \frac{\partial b^*_{\boldsymbol{k}}}{\partial n} - b^*_{\boldsymbol{k}} \frac{\partial \varphi_{\boldsymbol{k}_0}}{\partial n} \right\} \mathrm{d}S = \tag{4.34}$$

$$= \sum_{\boldsymbol{R}} \exp[i(\boldsymbol{k} - \boldsymbol{k}_0)\cdot\boldsymbol{R}](E - E_k) \int_{\text{cell}} b^*_{\boldsymbol{k}} \varphi_{\boldsymbol{k}_0} \mathrm{d}r .$$

From this equation it immediately follows that we need consider only those values of $\boldsymbol{k}$ which differ from $\boldsymbol{k}_0$ by a vector of the reciprocal lattice, i.e. $\boldsymbol{k} = \boldsymbol{k}_0 + \boldsymbol{K}$. We also rearrange the surface integral in the same manner as in Sect. 1, pairing off conjugate points on each cell boundary and using the fact that $\varphi_{\boldsymbol{k}_0}$ satisfies the exact boundary conditions. In this way we obtain

$$(4.35)\qquad \sum_{\boldsymbol{R}} \int_{\text{cell}} \frac{1}{2}\left\{\varphi_{\boldsymbol{k}_0}\left(\frac{\partial b_{\boldsymbol{k}}^*}{\partial n} + \frac{\partial b_{\boldsymbol{k}}^{*\prime}}{\partial n'}\right) - \frac{\partial \varphi_{\boldsymbol{k}_0}}{\partial n}\,(b_{\boldsymbol{k}}^* - b_{\boldsymbol{k}}^{*\prime})\right\} \mathrm{d}S = (E - E_{\boldsymbol{k}}) \int_{\text{cell}} b_{\boldsymbol{k}}^* \varphi_{\boldsymbol{k}_0}\, \mathrm{d}\boldsymbol{r}\,.$$

We use this condition and also assume that $\varphi_{\boldsymbol{k}_0}$ can be expressed as a series

$$(4.36)\qquad \varphi_{\boldsymbol{k}_0} = \sum_{\boldsymbol{K}'} a_{\boldsymbol{k}_0+\boldsymbol{K}'} b_{\boldsymbol{k}_0+\boldsymbol{K}'}\,.$$

This is the most general possible series of functions of the form (4.32) which can still satisfy the Bloch conditions corresponding to $\boldsymbol{k}_0$. Using (4.35) we then obtain from (4.34) the result:

$$(4.37)\qquad \sum_{\boldsymbol{K}'} a_{\boldsymbol{k}_0+\boldsymbol{K}'}(E - E_{\boldsymbol{k}_0+\boldsymbol{K}}) \int_{\text{cell}} b_{\boldsymbol{k}_0+\boldsymbol{K}}^* b_{\boldsymbol{k}_0+\boldsymbol{K}'}\, \mathrm{d}\boldsymbol{r} =$$
$$= \sum_{\boldsymbol{K}'} a_{\boldsymbol{k}_0+\boldsymbol{K}'} \int_{\text{cell}} \frac{1}{2}\left\{- b_{\boldsymbol{k}_0+\boldsymbol{K}'} \frac{\partial b_{\boldsymbol{k}_0+\boldsymbol{K}}^{*\prime}}{\partial n} + \frac{\partial b_{\boldsymbol{k}_0+\boldsymbol{K}'}}{\partial n}\, b_{\boldsymbol{k}_0+\boldsymbol{K}}^{*\prime}\right\} \mathrm{d}S\,.$$

If we now use the definition (4.32) of the $b_{\boldsymbol{k}}$'s we obtain, after some use of the symmetry properties of b's for crystals having a center of inversion, secular equation:

$$(4.38)\qquad \sum_{\boldsymbol{K}'} a_{\boldsymbol{k}_0+\boldsymbol{K}'}(E - E_{\boldsymbol{k}_0+\boldsymbol{K}}) \int_{\text{cell}} v_{\boldsymbol{s}_0+\boldsymbol{K}}^* v_{\boldsymbol{k}_0+\boldsymbol{K}'} \exp\left[i(\boldsymbol{K}' - \boldsymbol{K})\cdot\boldsymbol{r}\right] \mathrm{d}\boldsymbol{r} =$$
$$= \frac{1}{2} \int_{\text{cell}} \left\{v_{\boldsymbol{k}_0+\boldsymbol{K}}^{*\prime} \frac{\partial v_{\boldsymbol{k}_0+\boldsymbol{K}'}}{\partial n} - \frac{\partial v_{\boldsymbol{k}_0+\boldsymbol{K}}}{\partial n}\, v_{\boldsymbol{k}_0+\boldsymbol{K}'}\right\} \exp\left[i(\boldsymbol{K}' - \boldsymbol{K})\cdot\boldsymbol{r}\right] \mathrm{d}S +$$
$$+ i \int \hat{n}\cdot\left\{\boldsymbol{k}_0 + \frac{\boldsymbol{K} + \boldsymbol{K}'}{2}\right\} v_{\boldsymbol{k}_0+\boldsymbol{K}}^{*\prime} v_{\boldsymbol{k}_0+\boldsymbol{K}'} \exp\left[i(\boldsymbol{K}' - \boldsymbol{K})\cdot\boldsymbol{r}\right] \mathrm{d}S\,.$$

On examining the surface integrals which occur in (4.38) we note that they are quite similar in form to those which occur in the first order perturbation (4.9). In particular if the Bloch functions are free electron functions then the surface integrals vanish and the secular problem is diagonal, thus fulfilling the empty lattice test. If the wave functions are fairly close to those for free electrons the surface integrals will be small, and we are probably justified in ignoring the secular equation and considering only the first order perturbation. There will, however, be certain values of the wave vectors for

which two or more spherical approximation wave functions are degenerate, or nearly degenerate, with each other. This will occur whenever

$$|(\boldsymbol{k}_0 + \boldsymbol{K})| = |(\boldsymbol{k}_0 + \boldsymbol{K}')| \tag{4.39}$$

a condition which follows from the fact that in the spherical approximation the energy can depend only on the magnitude and not on the direction of $\boldsymbol{k}_0 + \boldsymbol{K}$. The condition (4.39) is satisfied whenever $\boldsymbol{k}_0 + \boldsymbol{K}$ is a vector terminating on the perpendicular bisector plane to the vector $\boldsymbol{K} - \boldsymbol{K}'$. This is just the usual diffraction condition for two wave vectors in X-ray, neutron, or electron diffraction. The solutions of the spherical problem are thus the natural generalizations of plane waves in the almost free electron theory of Brillouin [4]. It is the perturbations between functions of different wave vector—differing in fact by a vector of the reciprocal lattice—which give rise to energy gaps at the Brillouin Zone boundaries in $\boldsymbol{k}$ space. Near such planes, on which two different solutions of the spherical problem are degenerate, the true solution of the Bloch problem corresponds to an admixture of two or more solutions of the spherical problem, no one of which has predominant weight. In regions of $\boldsymbol{k}$ space away from planes of discontinuity one particular solution of the spherical problem predominates and the energy is then given to a good approximation by the solution of the spherical problem and probably to a good approximation in many cases by the power series expansion of the energy (3.30). At planes of discontinuity the boundary perturbations remove the degeneracy which exists in the spherical approximation. However, the sum of the energy levels arising at a given discontinuity plane should be very close to the energy given by the spherical approximation.

An alternative procedure to that described above would have been to start with plane waves and to treat the potential of the crystal as a perturbation. We should have then obtained a secular equation very similar to that arising from (4.38) except that the right hand side would be replaced by matrix elements of the potential between different plane waves, and on the left hand side the terms in the sum with $\boldsymbol{K} \neq \boldsymbol{K}'$ would have dropped out because plane waves are mutually orthogonal. However, this secular problem is poorly convergent, i.e. we would have had to retain many plane waves to obtain a good solution, and the perturbations would have been large even away from planes of discontinuity. If we orthogonalize the plane waves to the core states the convergence is much improved, that is the matrix elements of the crystal potential between orthogonalized plane waves are much smaller than the matrix elements between ordinary plane waves. In the procedure described in this chapter the matrix elements of the boundary perturbation between different solutions of the spherical problem are even smaller than the corresponding matrix elements between plane waves. In other words the so-

lutions to the spherical problem diagonalize the overall problem to an even better degree than do ordinary orthogonalized plane waves.

For the cohesive energy problem the main lesson of the above discussion is that we may be justified in using the spherical approximation to discuss sums over many k states even when individual levels are rather poorly described, especially near zone boundaries, by this approximation. It appears that the « off-diagonal » surface integrals in (4.38), i.e. those involving different k's for the two functions, should be of the same order of magnitude as the diagonal integrals which occurred in the first order boundary perturbation. If the boundary perturbations are small, then the energy levels will be affected significantly by these boundary perturbations only very close to Brillouin Zones in k-space. In such cases the individual levels may be shifted from their spherical values by a few hundredths of a rydberg unit, but only a small fraction of the energy levels in the whole Fermi distribution are so affected even in polyvalent metals like Mg or Al. Thus the effect on the cohesive energy should be extremely small even in many cases where the Fermi distribution includes more than one Brillouin Zone. When we come to discuss such matters as the electron transport properties, especially galvanomagnetic effects, the de Haas-van Alphen effect, etc., the properties are sensitive primarily to the energy levels of the electrons close to the Fermi surface, and hence are much more sensitive to the details of the energy level structure. Even the electronic specific heat and the free spin susceptibility depend on the energy levels at the Fermi surface and are much more sensitive to deviations from the spherical approximation than is the cohesive energy. Elastic constants, too, are sensitive to changes in shape of the energy surfaces away from spherical symmetry, and thus would depend on solutions of the secular equation derived from (4.38), i.e. the spherical approximation would not be adequate to deal with the problem, except for the pressure-volume relations.

It should be remarked that the boundary perturbation is not the only perturbation that must be taken into account when we come to discuss the details of the energy level structure of electrons. Small deviations of the potential from spherical symmetry are of equal or possibly greater importance. Also we note that as in the case of eq. (4.15), (4.26) and (4.28), the « matrix elements » occurring in (4.38) depend quite strongly on the exact details of the potential very close to the cell boundary. Thus I believe that calculations of details of the energy band structure are meaningless unless great care is taken in constructing the potential in the outer part of the cell. To my knowledge only one calculation which meets these requirements has appeared in the literature, that of V. Heine for Aluminium [5].

REFERENCES

[1] D. J. Howarth and H. Jones: *Proc. Phys. Soc.*, A **65**, 355 (1952).
[2] J. Bardeen: *Journ. Chem. Phys.*, **6**, 367 (1938).
[3] Much of the material in this lecture is new and has not yet been published. This applies particularly to the discussion of Sect. **5**.
[4] L. Brillouin: *Compt. Rend.*, **191**, 198, 292 (1930)
[5] V. Heine: *Proc. Roy. Soc.*, A **240**, 340 (1957).

V.

The quantum defect method.

1. – Introduction.

The discussion of Chapters III and IV has shown that in general all that we need to solve the one-electron part of the solid state problem is the following information:

1) The values of the logarithmic derivatives φ_l as a function of energy E at radii in the vicinity of the cell boundary.

2) Details of the cell potential in the immediate vicinity of the cell boundary.

Only for the calculation of the details of the band splitting near the BZ boundary is it necessary to know anything about the interior wave functions, and then they enter only into calculation of the non-orthogonality between different waves, and to some extent into the calculation of effects due to the anisotropy of the potential within the cell. These two effects, while of importance in the prediction of physical properties which depend sensitively on the details of the band structure, such as galvanomagnetic effects, or, in the case of polyvalent metals, elastic constants, nevertheless have very little influence on the cohesive energy, which depends only upon average energies taken over all the Fermi distribution.

If an explicit potential is available the most straightforward procedure is to numerically integrate the radial equations (3.17) and determine the φ_l's at the appropriate r_s values. The most accurate calculations using this procedure have been made for Li [1] and Na [2], and more recently K [3]. In the case of Li and Na potentials were constructed semi-empirically so as to give energy levels for the free atom which agreed with observed term energies

in the free atom spectra. For Na the potential was constructed according to a method given by PROKOFJEW [4] who used the WKB approximation. JOST and KOHN [5] have shown that the potential is uniquely determined by the eigenvalues provided eigenvalues for more than one angular momentum quantum number are used. Alternatively if the potential is assumed to be rigorously Coulombic or to have some other rigorous form outside a cut-off radius r_i, then again it can be shown by similar arguments to those of JOST and KOHN [6] that the potential is uniquely determined by the eigenvalues for one angular momentum. In this case, however, it is possible that different effective potentials pertain to different values of the angular momentum quantum number, and this fact will be true, especially for the heavier elements, as a result of exchange and correlation effects. PROKOFJEW, then, assumed a potential which was Coulombic outside the ion core, and determined the remainder of the potential from the condition that the Bohr-Sommerfeld quantization conditions

$$\varphi\left[E - V(r) - \frac{(l+\frac{1}{2})^2}{r^2}\right]^{\frac{1}{2}} \mathrm{d}r = \left(m - l - \frac{1}{2}\right)\pi\,, \tag{5.1}$$

where m is an integer, give the observed eigenvalues. In the case of Na it was possible to construct a single potential for all values of l which would reproduce the observed term values to better than 1 per cent.

In the case of Li similar principles were used but explicit radial integrations with the trial potential were used to adjust the potential to agree with the eigenvalues. In this case the observed spectrum for s, p and d terms could be reproduced to an accuracy of a few tenths of one per cent. The greater accuracy for Li probably reflects the fact that exchange and core polarization effects are less important and therefore a single potential can represent the eigenvalues for all l values with greater accuracy in the case of Li than Na.

In the case of K Hartree-Fock potentials determined for the free atom were used, but with a correction for the difference in exchange interaction with the core for s, p and d electrons [3]. Attempts were also made to take into account core polarization [7]. Hartree-Fock potentials with corrections for difference in exchange potential for s and p electrons have also been used by HEINE [8] for Al, and by HERMAN *et al.* [9] for Ge. The potential arising from Hartree-Fock charge densities has been used by various authors for Cu [8] and for Cs [9].

Coming back now to the cases of Li and Na, we note that the route by which we arrive at the surface properties of the logarithmic derivatives is a rather circuitous one, considering that the only information actually used is the free atom term scheme and the assumption that the potential is Coulombic outside the ion core. Because of this Van Vleck felt that there must be a more

direct route of arriving at the same information without explicit construction of the potential. Two methods which accomplished this result were worked out in detail by KUHN and VAN VLECK [10] and by KUHN [11]. In the first method an attempt was made to extrapolate the logarithmic derivatives directly as a function of energy. That is, values of logarithmic derivatives were computed at a radius R for several observed eigenvalues of the free atom and for a given value of l. It was found that if R were chosen sufficiently small but still outside the core, the logarithmic derivatives could be extrapolated smoothly as a function of energy down to the energy of the lowest valence electron state in the solid. The linear combination of Coulomb wave functions which matched the logarithmic derivative at R at a given energy could then be determined directly, and in this way the logarithmic derivative at a larger radius r_s could be found. This method worked satisfactorily for the s functions, but failed to produce satisfactory results for the p functions and the effective mass for reasons which are still not entirely clear.

In the Kuhn method [11] a different quantity was extrapolated. In one-electron spectra, such as those of alkali atoms, the eigenvalues may be expressed in the form

$$E_{l,m} = -\frac{1}{(m-\delta_l)^2}\,. \tag{5.2}$$

For a given value of l the quantity δ_l changes only slowly as we move from one term to the next in a series of energy levels belonging to the same value of l. Thus it is possible to obtain δ_l as a smooth function of the energy E. Thus the quantity $\delta_0(E)$ proves to be one which can be extrapolated graphically as a function of energy fairly readily to energies appropriate to the solid. KUHN was further able to relate the quantity δ_0 to the properties of solutions of the radial differential equation in the following manner. We consider first the s radial equation in the form

$$\frac{\mathrm{d}^2 f_0}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}f_0}{\mathrm{d}r} + (E-V)f_0 = 0\,. \tag{5.3}$$

Next we introduce the following changes of variable:

$$x = ln(-Er) = ln\frac{r}{n^2} \tag{5.4a}$$

$$u(x) = \exp\left[-\frac{x}{2}\right] r f_0(r)\,. \tag{5.4b}$$

The differential equation (5.3) then takes the form

$$\frac{\mathrm{d}u^2}{\mathrm{d}x^2} + P(x)u = 0\,, \tag{5.5}$$

where

$$P(x) = \left[- n^2 e^{2x} - n^4 e^{2x} V(n^2 e^x) - \frac{1}{4} \right]. \tag{5.6}$$

The equation (5.5) has two « turning points » at which $P(x) = 0$. The inner turning point, corresponding to small values of r or x, always occurs close to the nucleus where $V(r)$ is non-Coulombic, or at least does not correspond to a singly charged ion, but the outer turning point occurs in the Coulombic region where $V(r) = -2/r$. Thus the outer turning point is given by

$$\frac{r}{n^2} = 1 + \sqrt{1 - \frac{1}{4n^2}}. \tag{5.7}$$

For example, for Na, for which $1/n^2 = .601$, the turning point occurs at $r = 3.2$, which is not too much smaller than the observed radius of 4.0. However, it is also easily shown that the eigenvalue of (5.3) corresponding to the boundary condition $f_0'(E, r_s) = 0$ has a minimum when $r/n^2 = 2$, or $1/n^2 = 2/r_0$, where r_0 is the radius for which the minimum occurs. This relation will be discussed further in Chapter VI. This radius r_0 occurs very close to the outer turning point of (5.5). Explicitly we have to a good approximation $r = r_0 - \frac{1}{8}$, so that the turning point is just inside r_0. For Na it turns out that $r_0 = 2.99$, so that the turning point differs from r_0 only by about 4 per cent. Imai [12] has developed an approximate method of solution of equations like (5.5) which is an extension of the so-called WKB method. We shall not go into details but only note that Imai's method gives the solution in the form of a power series expansion in x about the turning point which has a high accuracy in the immediate neighborhood of the turning point. Now, because r_0 is so close to the turning point Kuhn was able to use Imai's method of solution to obtain an accurate expression for $f_0'(E, r_s)$ which involved only the quantum defect extrapolated to energy E, as a parameter. In this way Kuhn found it possible to adjust E to satisfy the boundary condition. Kuhn's result in fact gave the minimum ground state energy $E_0 = 2/r_0$ as a single universal function of the quantum defect δ_0. Thus, it was possible to plot both $\delta_0(E)$ from the relation (5.2) and from Kuhn's result on the same graph for a given metal. The energy at which the two curves intersected was then the minimum energy of the lowest valence state, and the corresponding value of r_0 was immediately determined.

Later we shall discuss Kuhn's procedure in more detail, but using a method capable of greater accuracy.

Unfortunately, it is not possible to generalize Kuhn's result for the purpose of computing the logarithmic derivatives for higher l. Suppose, for example, we wished to consider $l = 1$, also at the minimum energy. For

$l = 1$ the equation corresponding to (5.7) takes the form

$$\frac{r}{n^2} = 1 + \sqrt{1 - \frac{(l + \frac{1}{2})^2}{n^2}} . \tag{5.8}$$

and using the numerical values appropriate to Na we are immediately faced with the fact that (5.8) has no real solution. This does not mean there is not a turning point, but only that the turning point lies inside the atomic core. Thus the whole basis of the Kuhn method breaks down, and the situation becomes worse as we go to higher angular momenta. It is evident that a more powerful technique must be developed if we are to pursue the philosophy of the Kuhn-Van Vleck method.

2. – Extension of the quantum defect method.

Let us again return to a consideration of the radial differential equation (3.17). Outside the ion core we will take $V(r) = -2/r$. If this is done equation (3.17) becomes one of the standard forms of the confluent hypergeometric differential equation. Being a second order differential equation it has two linearly independent solutions. A general solution may therefore be written in the form

$$f_l(E, r) = \alpha_l(n) F_l(n, r) + \gamma_l(n) G_l(n, r) . \tag{5.9}$$

Here $F_l(n, r)$ and $G_l(n, r)$ are two linearly independent solutions of (5.3) such that F_l is regular at $r = 0$ and behaves as r^l and G_l is irregular at $r = 0$ and behaves as r^{-l-1}. The particular form of solution used for general arguments is not important, but in most practical calculations by the Q.D. method the solutions F_l and G_l are related to the forms defined by WANNIER [13]. To assist the reader in following the mathematical literature we give the exact relations in (5.10) below:

$$\begin{cases} rF_l(n, r) = \dfrac{z}{2} J^n_{2l+1}(z) ; \qquad rG_l(n, r) = \dfrac{z}{2} N^n_{2l+1}(z) \\ z = (8r)^{\frac{1}{2}} ; \qquad E = -\dfrac{1}{n^2} . \end{cases} \tag{5.10}$$

The properties of the functions in (5.10) have been discussed by KUHN [11] and by HAM [15]. The latter has also prepared extensive numerical tables [15]. More recently, more extensive tables have been compiled by HANKIN [16] who numerically integrated (5.3) using Ham's solution for small r as initial conditions.

Returning now to equation (5.9) we find that for an eigenvalue of the free atom the following relationship holds:

$$\frac{\gamma_l(n)}{\alpha_l(n)} = \frac{n^{2l+1}\Gamma(n-l)}{\Gamma(n+l+1)} \operatorname{tg} \pi(n-l-1) . \tag{5.11}$$

The relation follows simply from the fact that at an eigenvalue the solution (5.9) must be the one which vanishes at $r = \infty$. By using (5.2) we note that (5.11) can be written in the equivalent form

$$\frac{\gamma_l(n)}{\alpha_l(n)} = -\frac{n^{2l+1}\Gamma(n-l)}{\Gamma(n+l+1)} \operatorname{tg} \pi\delta_l(n) . \tag{5.12}$$

Although (5.12) is valid only at free atom eigenvalues the idea suggests itself that it may also be valid for intermediate values of n or E. This, in fact, can be proved by arguments using the WKB method subject only to the restriction that the outer turning point of the differential equation (3.17) lies at a value of r for which the potential is rigorously of the Coulombic form $-2/r$. For the special case $l = 0$ (5.12) is just

$$\frac{\gamma_0(n)}{\alpha_0(n)} = -\operatorname{tg} \pi\delta_0(n) \tag{5.13}$$

and we find from (5.7) for the alkali metals that even for values of n corresponding to the solid state energy the turning point does lie in the Coulomb region. Thus we are justified in using the extrapolated quantum defect directly for determining the correct linear combination of Coulomb functions in accordance with (5.9). The WKB argument does not justify (5.13) rigorously; in fact, it only shows that (5.13) is valid between and beyond eigenvalues to within the same degree of accuracy that the Bohr-Sommerfeld quantization condition (5.1) gives the eigenvalues of the radial equation. However, the additional fact that (5.13) is rigorously true at eigenvalues lends almost overwhelming support to the use of (5.13).

When we come to higher l values, however, we find ourselves in trouble just as KUHN did because, as we have seen, the differential equation has no turning point in the Coulomb region for energies in the solid state range. The WKB type of argument can still be used to justify (5.12) *between* the free atom eigenvalues, but not in the region below these eigenvalues where we are interested. We can in fact see quite clearly that (5.12) cannot be correct in general for any extrapolated quantum defect. For example, for p states we may write

$$\frac{\gamma_1(n)}{\alpha_1(n)} = -\frac{\operatorname{tg} \pi\delta_1(\varepsilon)}{1-\varepsilon} , \qquad \varepsilon = -E = \frac{1}{n^2} . \tag{5.14}$$

Equation (5.14) has a singularity at $\varepsilon = 1$ unless $\delta_1(\varepsilon)$ becomes an integer for the same energy. There is, of course, no guarantee that the $\delta_1(\varepsilon)$ extrapolated from the observed eigenvalues will go through such a point in general. In order to extrapolate through the « forced zero » the p quantum defect curve will have to undergo more or less rapid variation. This rapid variation will begin for the energy for which Eq. (5.8) no longer has a real root, i.e. for $\varepsilon > .44$ Ry.

For the d levels the relation corresponding to (5.14) becomes

$$\frac{\gamma_2(n)}{\alpha_2(n)} = -\frac{\operatorname{tg} \pi \delta_2(\varepsilon)}{(1-\varepsilon)(1-4\varepsilon)}\,, \tag{5.15}$$

which shows that $\delta_2(\varepsilon)$ must have forced zeroes at $\varepsilon = 1$ and $\varepsilon = .25$. Thus the δ_2 extrapolation is even more distorted than the δ_1 and begins at $\varepsilon = .16$ Ry.

On the other hand, it can be shown that $\gamma_l(\varepsilon)$ is an analytic function of ε with no singularities in the finite complex plane of ε, and it can therefore be expanded in a power series in ε which converges everywhere in the finite plane. The other quantity $\alpha_l(\varepsilon)$ can be expressed in the form

$$\alpha_l(\varepsilon) = \beta_l(\varepsilon) - \gamma_l(\varepsilon) G_l(l, \varepsilon)\,, \tag{5.16}$$

where $\beta_l(\varepsilon)$ and $\gamma_l(\varepsilon)$ are both analytic throughout the plane and $G_l(l, \varepsilon)$ is a fixed function which is very small and smooth on the real axis of ε. By such arguments we are led to the expectation that the quantity defined by

$$\frac{\gamma_l(n)}{\alpha_l(n)} = -\operatorname{tg} \pi \eta_l(\varepsilon)\,, \tag{5.17}$$

should be a function of ε which can be smoothly extrapolated. This new function $\eta_l(\varepsilon)$, known henceforth as the η-defect, is free of poles on the real axis of ε, and has no forced zeroes.

Ham [17] has tested the η-defect extrapolation by actually solving the differential equations (3.17) using the following test potential:

$$\begin{cases} V(r) = -\dfrac{2z}{r} & r < r_0\,, \\ V(r) = -\dfrac{2}{r} & r > r_0\,, \\ z = 1.44 & r_0 = 3.125\,. \end{cases} \tag{5.18}$$

Plots of $\delta_l(\varepsilon)$ and $\eta_l(\varepsilon)$ for this potential are shown in Fig. 3 *a*, *b*, *c* for $l = 1, 2, 3$, respectively.

These figures well illustrate all the points mentioned in the text. HAM has tested the accuracy with which the $l = 1$ η-defect can be extrapolated by means of quadratic and cubic polynomials in ε. Typical results are shown in Table XIX. The exact results refer to the potential (5.18) and they were matched at the ε with zero error in the table.

In practice for the alkali metals it develops that the differences between the η-defect and δ-defect extrapolations are not large. In the case of the monovalent noble metals the difference is greater. In both cases, however, η is a more slowly varying function of energy than δ. Another trick which has proved helpful is to make use of core energy levels. HAM [18] shows that it is the ionization potentials of the already singly ionized ions which should be used. Not very great accuracy is needed in these eigenvalues as the sensitivity of the quantum defect to the energy for very large binding energies above 1 Ry is very small. This extra information has proved of special importance in the case of Cu, Ag and Au, where in some cases the ground state energy of the valence electron lies beyond 1 Ry and hence beyond the forced zero of the p defect [19].

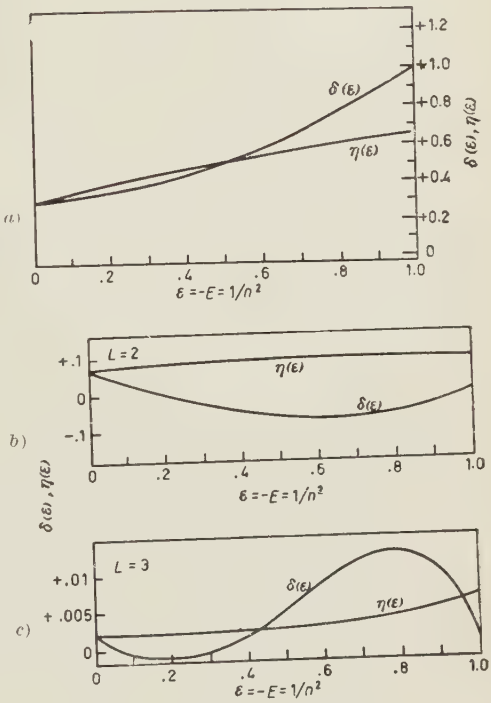


Fig. 3. – Plots of $\delta(\varepsilon)$ and $\eta(\varepsilon)$ for the potential (5.17) as computed by HAM.

TABLE XIX.

ε	Exact η_1	$\Delta\eta_1$-quadratic	$\Delta\eta_1$-cubic
0.1	.33254	0	0
0.2	.37825	0	0
0.3	.42342	0	0
0.4	.46728	+.00077	—
0.5	.50786	+.00428	+.00027
0.6	.54383	+.01186	+.00084
1.0	.64034	+.08415	0

3. – The problem of d-states.

When we come to discuss the d-defect even the η-defect extrapolation proves to be inadequate. For Li and Na the quantum defect for $l = 2$ may be taken as 0, so there is no problem, but starting with K the defect departs appreciably from 0. For K the lowest d level is at 0.123 Ry, and we are required to extrapolate to 0.48 Ry for the ground state energy. By contrast the lowest p level is at 0.2 Ry and in addition we have a core state at 2.34 Ry which gives an η-defect differing very little from that in the free atom spectrum range. In the case of Rb and Cs we have a similar situation; in addition we have core d-states whose η-defect differs quite substantially from the values in the free atom term value range. The situation is well illustrated by Table XX, which shows the experimental η-defects for p and d-states in the case of Cs.

TABLE XX.

Rb			Cs		
m	p	d	m	p	d
3		(0.999)	4		(1.996)
4	(2.706) (*)	1.3624	5	(3.710)	2.4828
5	2.6798	1.3530	6	3.6258	2.4791
6	2.6616	1.3504	7	3.5961	2.4758
7	2.6551	1.3493	8	3.5852	2.4730

(*) Core states enclosed in ( ).

There is a further and more fundamental difficulty with d-states, however, which is associated with the normalization of the wave function. It will be noted that the normalization theorem (3.44) gives us

$$\frac{\partial}{\partial \varepsilon} \frac{r_s f_l'(r_s)}{f_l(r_s)} = \varphi_l' = \int_0^{r_s} \frac{f_l^2(r)}{f_l^2(r_s)} \frac{r^2 \, \mathrm{d}r}{r_s} . \tag{5.19}$$

Since the integrand is inherently positive it follows that the quantity φ_l' must also be positive. This requirement sets restrictions on the possible values and variation of the η-defect. For s and p-states, it is found in practice that the logarithmic derivatives of both the irregular and regular radial functions are increasing functions of ε. As a result the requirement that the actual φ_l' be positive does not impose any restriction on the *value* of η, and sets only a mild

upper limit on the possible increase of η with ε. For the d Coulomb functions, however, we find that for the energies encountered in the solid the logarithmic derivative of the irregular function is decreasing with ε. This means that the wave function can never consist entirely of irregular function, and therefore that the extrapolated η_2-defect can never cross a half-integer value for any value of the energy in the neighborhood of the valence ground state energy—or more precisely any energy for which the logarithmic derivative of the irregular function decreases with increasing ε. For Rb and Cs it is found that a simple quadratic interpolation formula fitted to the observed term values extrapolates across a half integer, and hence cannot be a correct extrapolation. HAM has computed the actual variation of η_2 with energy for d-states in a square well potential having about the strength of the actual potential (in terms of the lowest d-eigenvalues), and by requiring the actual η_2-defect curve to be similar in shape to the curve for the artificial potential while at the same time conforming to the observed eigenvalues and satisfying the requirements for a positive φ_2; HAM was able to set reasonable upper and lower limits to the η_2-defect curve.

As an illustration of the application of these results we show in Table XXI values for the upper and lower limits of η_2 for Cs obtained for the ground state energies corresponding to various values of the sphere radius r_s. The Table

TABLE XXI.

	η_2		$(5/7)(E_4/E_2)k_0^2$	
r_s	Upper	Lower	Upper	Lower
4.6	.371	.317	— .330	— .153
5.0	.376	.323	— .1727	— .093
5.4	.381	.335	— .1113	— .0691
5.8	.388	.347	— .0831	— .0578

also shows the corresponding limits for the ratio $(5/7)(E_4/E_2)k_0^2$, which is the ratio of the contribution of the k^4 term in the expansion of E to the k^2 term to the mean kinetic energy of the valence electrons.

It should be noted in passing that the difficulty in the extrapolation of the η_2-defect is to some extent an artificial one, which was introduced by our effort to obtain the energy as a power series in k^2. In Chapter III we pointed out that it was possible to solve the determinant of (3.31) either directly or by expanding all the elements in powers of k and then collecting terms in the same power of k in the final expansion of the determinant. The direct method is probably most advantageous when there exist uncertainties in the extra-

polation of the η-defects, since in this method a knowledge of the wave functions only at the actual solid state energy and not the ground state energy is needed. Thus for the Fermi level of Cs, which lies at approximately — .285 Ry, the extrapolation of the η_2-defect is very much less severe than for the ground state energy of — .42 Ry. A comparison of the direct and expansion methods for Cs at the Fermi level is shown in Table XXII, where again the limiting

TABLE XXII. – *Fermi level of* Cs ($r_s = 5.6$).

Power Series	— .2891	— .2818
Determinant	— .2837	— .2834

values from Ham's estimated η_2-extrapolation are shown. We see from this Table that the upper and lower limits for the position of the Fermi level are much closer for the determinantal evaluation, and indeed the positions are reversed, since the Table shows the levels associated with a given η_2 curve. More generally, we can say that the farther out we go in k-space in the spherical approximation, the more sensitive does the energy become to the accuracy of extrapolation of the η_2-defect, but if the direct expansion of (3.31) is used it also follows that the larger the k the more nearly are we in the free atom eigenvalue range, and hence the more reliable the extrapolation or interpolation procedures.

4. – Corrections for changes in potential.

A basic limitation of the quantum defect method (Q.D.M.) as so far described is that the potential in the atomic polyhedron must be assumed to be strictly Coulombic. A further assumption is that the potential from which the experimental free atom eigenvalues are derived can also be considered as strictly Coulombic outside some given radius. Neither of these assumptions can be strictly fulfilled in practice. We have already seen that the self-consistent potential —even when exchange and correlation effects are properly taken into account— cannot be strictly Coulombic in the outer parts of the cell. Furthermore, in heavy atoms especially, it is well know that there is a component of the potential, due to the polarizability of the ion core, which falls off as r^{-4} and which is large enough to have a non-negligible effect on the energy levels of the free atom. This so-called polarization potential has two difficult aspects. In the first place, its presence, strictly speaking, invalidates the whole argument by which the logarithmic derivatives of the wave functions were obtained from the quantum defects. In the second place, the polarization potential will be present in the solid and will be an additional deviation from the strictly Coulombic assumption.

The deviation from Coulomb potential in the solid is the simpler of the two problems and we shall deal with it first.

We let $V_s(r)$ represent the standard potential, which we will take as Coulombic $-2/r$, and we let $V_p(r)$ represent the true potential in the atomic sphere in the spherical approximation. We let $f_l(r)$ represent the correct radial solution and $g_l(r)$ the solution for the standard potential, so that we may write:

$$f_l'' + \frac{2}{r} f_l' + \left[E - V_s(r) - \frac{l(l+1)}{r^2}\right] f_l = 0\,, \tag{5.20a}$$

$$g_l'' + \frac{2}{r} g_l' + \left[E - V_s(r) - \frac{l(l+1)}{r^2}\right] g_l = 0\,. \tag{5.20b}$$

We subtract these equations and integrate from 0 to r_s assuming that $f_l(r)$ and $g_l(r)$ are both properly behaved at $r = 0$ for energy E. We may also assume that the logarithmic derivatives of $f_l(r)$ and $g_l(r)$ are identical at the origin, since this involves only the assumption that the origin is a regular singular point of the differential equations (5.20). We then obtain

$$\frac{f_l'(r_s)}{f_l(r_s)} - \frac{g_l'(r_s)}{g_l(r_s)} = \int_0^{r_s} [V_p(r) - V_s(r)] \frac{f_l(r)}{f_l(r_s)} \frac{g_l(r)}{g_l(r_s)} \frac{r^2}{r_s^2}\, dr\,. \tag{5.21}$$

This relation is exact, but we may obtain an approximate relation, equivalent to first order perturbation theory, by taking $f_l(r) = g_l(r)$ in the integral. Bearing in mind the definition of φ_l, and assuming $g_l(r_s) = 1$ for all E, we obtain:

$$\Delta\varphi_l = \int_0^{r_s} \Delta V(r) \frac{r^2}{r_s} g_l^2(r)\, dr\,, \tag{5.22}$$

where $\Delta\varphi_l$ is the change in φ_l resulting from the change in potential $\Delta V(r)$. The relation (5.22) is valid for all E or ε and may therefore be differentiated with respect to E as the occasion requires.

In making calculations with (5.22) a useful approximation is the following. Since we are interested mainly in alterations in potential in the outer part of the cell, we shall use an approximation which gives a good representation of the behavior of $g_l(r)$ in the immediate vicinity of r_s, but may not be so good elsewhere. We assume that $f_l(r)$ can be expanded in a Taylor series about $r = r_s$. If we retain terms up to the second order in the expansion we obtain

$$g_l^2(r) = 1 - 2q_l\left(1 - \frac{r}{r_s}\right) + [q_l^2 - 2q_l + l(l+1) + r_s^2(V - E_0)]\left(1 - \frac{r}{r_s}\right)^2. \tag{5.23}$$

Using this result in (5.22) we obtain:

$$\Delta\varphi_l = \Delta V_1 - 2\varphi_l \Delta V_2 + [\varphi_l^2 - 2\varphi_l + l(l+1) + r_s^2(V - E_0)]\Delta V_3 , \tag{5.24}$$

where

$$\Delta V_1 = \frac{1}{r_s}\int_0^{r_s} \Delta V(r) r^2 \, dr , \tag{5.25a}$$

$$\Delta V_2 = \frac{1}{r_s}\int_0^{r_s} \Delta V(r) \left(1 - \frac{r}{r_s}\right) r^2 \, dr , \tag{5.25b}$$

$$\Delta V_3 = \frac{1}{r_s}\int_0^{r_s} \Delta V(r) \left(1 - \frac{r}{r_s}\right)^2 r^2 dr . \tag{5.25c}$$

This equation is only valid if $V(r)$ is assumed to vanish inside a radius r_i for which the expansion (5.23) becomes inaccurate. Of course, since we know the wave functions exactly in the region of Coulombic potential we could evaluate (5.22) without using the approximation (5.23), but the greater accuracy is probably not worth while in most cases, and the form (5.24) has the virtue of bringing out the physical principles involved more clearly.

Suppose we assume that $\Delta V(r)$ vanishes for $r < r_i$, but that we add to it a constant potential C which extends throughout the cell, and which is adjusted to make ΔE_0 vanish. This constant potential then represents minus the energy shift of the ground state energy resulting from the perturbation in first order. In this way we obtain the following results

$$C = -\frac{\Delta V_1}{\varphi_0'} - \frac{r_s^2(V - E_0)}{\varphi_0'} \Delta V_3 , \tag{5.26}$$

$$\frac{\Delta\varphi_l}{\varphi_l} = \Delta V_1 \frac{1}{\varphi_l}\left(1 - \frac{\varphi_l'}{\varphi_0'}\right) - 2\Delta V_2 + \Delta V_3 \left[r_s^2(V - E_0) \frac{1}{\varphi_l}\left(1 - \frac{\varphi_l'}{\varphi_0'}\right) + \varphi_l + \frac{l(l+1)}{\varphi_l} - 2 \right] . \tag{5.27}$$

In obtaining these results we have used the relation (5.19) in evaluating the C term arising in (5.22). By taking derivatives of (5.27) with respect to ε, and taking into account the boundary conditions satisfied by φ_0, we obtain

the relations

$$(5.28)\qquad \frac{\Delta\varphi_0'}{\varphi_0'} = -\Delta V_1 \frac{\varphi_0''}{\varphi_0'^2} - 2\Delta V_2 + \Delta V_3\left[-r_s^2(V-E_0)\frac{\varphi_0''}{\varphi_0'^2} + \frac{r_s^2}{\varphi_0'} - 2\right],$$

$$(5.29)\qquad \frac{\Delta\varphi_0''}{\varphi_0''} = -\Delta V_1 \frac{\varphi_0'''}{\varphi_0''\varphi_0'} - 2\Delta V_2 + \Delta V_3\left[-r_s^2(V-E_0)\frac{\varphi_0'''}{\varphi_0''\varphi_0'} + 2\left(\frac{\varphi_0'^2}{\varphi_0''} - 1\right)\right],$$

$$(5.30)\qquad \frac{\Delta\varphi_1'}{\varphi_1'} = -\Delta V_1 \frac{\varphi_1''}{\varphi_0'\varphi_1'} - 2\Delta V_2 + \Delta V_3\left[-r_s^2(V-E_0)\frac{\varphi_1''}{\varphi_0'\varphi_1'} + \frac{r_s^2}{\varphi_1'} + 2(\varphi_1 - 1)\right].$$

From these relations we may compute the changes resulting from the perturbation of all the quantities that enter into the computation of E_2 and E_4 in equations (3.35) and (3.36). If we use the free electron values for the φ's and their derivatives, as given for example in (3.38), we immediately verify that the change in E_2 and E_4 vanishes for an arbitrary perturbation. Thus in this case the whole band structure is shifted rigidly by the amount C given in eq. (5.26). This result is not surprising when we consider that for free electrons the charge density in the cell is constant. It is true, however, only in the approximation of first order perturbation theory.

More generally we find that any potential perturbation which tends to flatten out the potential variation in the outer part of the cell, will tend to make the values of E_2 and E_4 approach closer to the free electron results.

As an illustration of these considerations we may consider the effect of the potential due to the other valence electrons. In order to avoid extending this into the core, we assume the following modified potential:

$$(5.31)\qquad \Delta V(r) = -\frac{\gamma r^2}{r_s^3}.$$

After we have determined the C due to this potential, we must also add a constant C_0 to the energy which makes the total potential vanish at $r = r_s$. This has the value $C_0 = 2 + \gamma/r_s$, where γ is defined by (3.45) and is usually very close to unity. As a first illustration we will consider the case of Li at $r_s = 3.2$. We shall somewhat arbitrarily cut off the potential (5.51) at $r_i/r_s = 0.5$, but it is found that the answers are quite insensitive to the choice of cut-off, since so little contribution to the averages (5.25) comes from the inner part of the cell. Quite generally for the potential (5.31) we find:

$$(5.32)\qquad \Delta V_1 = -.1939\gamma r_s, \qquad \Delta V_2 = -.02969\gamma r_s, \qquad \Delta V_3 = -.00736\gamma r_s.$$

For the Li case we find that

$$(5.33)\qquad C_0 - C = .7464,$$

while E_2 is raised from .73 to .76. We find that E_4 is changed from $-.03$ to $-.02$, an entirely negligible shift.

In computing the influence of these results on the cohesive energy we must remember that the effect of the potential of the other valence electrons on the electronic energy is really the Coulomb interaction of the electrons. In the one electron scheme it is counted twice. Thus the total correction to the energy is actually half the correction to the one electron energy less the Coulomb self energy computed in the free electron approximation, which is $1.2/r_s$. In this way we find a correction of $-.0018$ from the ground state energy and .0032 from the Fermi energy, resulting in a net decrease in cohesive energy of .0014 Ry or only 0.46 kcal per mol.

The influence of the potential (5.31) on the band structure can be quite considerable. For example, for Cs with $r_s = 5.4$ E_2 is reduced from 1.286 to 1.141, much closer to the free electron value. If we neglect changes in E_4, we find the total correction to the cohesive energy amounts to .0122, an increase of 3.8 kcal per mol. This illustrates the general point. For $r_s = 5.8$ the correction is .0083 Ry or 2.6 kcal per mol, in a direction to decrease the equilibrium lattice constant.

There will be a slight additional correction to the energies arising from second order perturbations of the potential (5.31). This correction is very small and therefore will not be discussed in detail here.

REFERENCES

[1] F. Seitz: *Phys. Rev.*, **47**, 400 (1935); the tabulated potential in this reference (Table II, p. 402) is incorrect, although calculations were carried out with the correct potential. For a tabulation of the correct potential see W. Kohn and N. Rostoker: *Phys. Rev.*, **94**, 1111 (1954).

[2] E. Wigner and F. Seitz: *Phys. Rev.*, **43**, 804 (1933); **46**, 509 (1934). This potential has been used for Na by almost all subsequent authors.

[3] E. Gorin: *Phys. Zeits. der Sowietunion*, **9**, 328 (1936); J. Callaway: *Phys. Rev.*, **103**, 1219 (1956); S. Berman, J. Callaway and R. D. Woods: *Phys. Rev.*, **101**, 1467 (1956).

[4] W. Prokofjew: *Zeits. f. Phys.*, **58**, 255 (1929).

[5] R. Jost and W. Kohn: *Phys. Rev.*, **87**, 977 (1952); **88**, 382 (1952).

[6] F. S. Ham: *Solid State Physics*, vol. I (New York, 1955), cfr. Theorem II, pp. 147-148.

[7] J. Callaway: Tech. Report No. 3, NR 817-616, University of Miami (unpublished).

[8] V. Heine: *Proc. Roy. Soc.*, **240** A, 340 (1957): *Thesis* (Cambridge University, 1956, unpublished).

[8a] D. J. Howarth: *Proc. Roy. Soc.*, A **220**, 513 (1953).

[9] F. Herman: *Phys. Rev.*, **95**, 847 (1954).

[9a] R. Sternheimer: *Phys. Rev.*, **78**, 238 (1950).
[10] T. S. Kuhn and J. H. Van Vleck: *Phys. Rev.*, **79**, 382 (1950).
[11] T. S. Kuhn: *Phys. Rev.*, **79**, 515 (1950).
[12] I. Imai: *Phys. Rev.*, **74**, 113 (1948).
[13] G. H. Wannier: *Phys. Rev.*, **64**, 358 (1943).
[14] T. S. Kuhn: *Q. Appl. Math.*, **9**, 1 (1951).
[15] F. S. Ham: Tech. Report No. 204, Cruft Laboratory, Harvard University.
[16] P. Hankin: unpublished.
[17] F. S. Ham: ref. [6], pp. 159-161.
[18] F. S. Ham: ref. [6], p. 165.
[19] K. Kambe: *Phys. Rev.*, **99**, 419 (1955).

VI.

The polarization correction.

1. – Inaccuracy of the quantum defects.

We have seen that the true potential seen by a valence electron in the free atom outside the ion core is given by

$$V(r) = -\frac{2}{r} - \frac{2\alpha'}{r^4} . \tag{6.1}$$

Thus the observed energy levels of the free atom include the effect of the extra polarization potential $-2\alpha'/r^4$, whereas in deducing the properties of the φ's from the quantum defects in the last lecture we assumed that the polarization potential was not present in (6.1). For strict applicability of the Q.D. arguments we should use the quantum defects not of a real atom but of a fictitious atom whose potential is identical with that of the real atom inside some radius r_i, but which is strictly Coulombic outside. If we chose r_i fairly large, however, the difference between the two potentials is quite small, and so we may use first order perturbation theory in reverse, as it were, to compute the energy levels of the fictitious atom from those of the real atom. The resulting corrections in the η-defects for all the alkalis have been tabulated by Ham [1]. The corrections were actually computed by an approximate method, and more recent unpublished calculations of Ham indicate that while the polarization corrections for $l = 0$ are about right, the values for $l = 1$ are considerably in error, and may be almost twice as large as shown in Ham's tables.

The corrections to the η-defects are negative and have the effect of raising the ground state energy and increasing the value of E_2 when the solid state parameters are computed from them.

2. – Polarization potential in the solid.

If we compute the solid state parameters with the revised quantum defects as described above, the results will, of course, be quite sensitive to the choice of cut-off radius r_i. The corrections to the energy will increase as r_i is decreased. The resulting energy will be that of a solid made up of fictitious atoms. In order to find the true solid state parameters we must estimate the effect of the difference in potential between the true and fictitious atoms according to the general principles described in Sect. **4** of Chapter V.

Our first inclination might be to take the polarization potential of the atoms in the solid the same as that for the free atoms. In this case it would only be necessary to take as $\Delta V(r)$ in (5.22)

$$\Delta V(r) = -\frac{2\alpha'}{r^4} \qquad r_i < r < r_s \tag{6.2a}$$

and zero elsewhere. This is valid if the Wigner-Seitz sphere radius lies outside the cut-off. On the other hand, if the sphere radius is inside the cut-off we must take the potential with the opposite sign, in accordance with (6.2*b*)

$$\Delta V(r) = +\frac{2\alpha'}{r^4} \qquad r_s < r < r_i \,. \tag{6.2b}$$

Simple physical considerations make it fairly apparent that the above hypothesis concerning the polarization potential is unrealistic. If the polarization of each atom extended to infinity, we would not be justified in neglecting the influence of the ion core in one cell on the electron in another. The argument developed in Chapter II for neglecting the electrostatic interactions from one cell to another does not apply to polarization because the polarization potential results from an induced dipole moment in the electronic distribution whose potential is not screened by a uniform charge distribution. On the other hand, if the polarization potentials did extend significantly outside the cells, the resulting contribution to the cohesion would be considerable. For Cs it would be of the order of .05 Ry, or nearly 15 kcal/mol, which is nearly equal to the experimental cohesive energy. This clearly is not a correct physical picture because we may expect that the valence electrons will in fact partially screen the polarization interaction with distant cores.

It is difficult to estimate the screening radius, and yet the final result is very sensitive to what this is. Two possible choices suggest themselves.

1) PINES [2] has suggested that the interaction between two electrons in a Fermi gas is screened with a radius given by

$$k_c^{-1} = 1.3 r_s^{\frac{1}{2}} \tag{6.3}$$

for the interaction between a valence electron and an ion core this radius should be halved because the screening occurs twice. The field exerted by the electron at the ion core is screened, so that the induced dipole is reduced by the screening factor. Simultaneously the field exerted by the dipole back at the electron is also screened. The resulting potential would thus take the form

$$V(r) = -\int_r^\infty \frac{10\alpha'}{r^5} \exp\left[-\frac{r}{.65 r_s^{\frac{1}{2}}}\right] \mathrm{d}r\,. \tag{6.3a}$$

The closest approach of a neighboring ion to the Wigner-Seitz sphere is $.756 r_s$, which is $1.16 r_s^{\frac{1}{2}}$ screening lengths, or about 2.7 screening lengths for Cs.

2) We may use the Thomas-Fermi screening length which gives a value of $0.322 r_s^{\frac{1}{2}}$, about $\frac{1}{2}$ that of case 1. Thus the nearest ion is about 5.4 screening lengths from the surface of the Wigner-Seitz sphere in the case of Cs. We may conclude that in this case the ion core polarization exerts a negligible effect outside its own cell.

Intermediate choices of length may be justified by some arguments, but the above represents the extreme limits. It seems probable that for this type of interaction the Fermi-Thomas screening length is more appropriate. If this is so, it means that not only is there no polarization potential outside the cell, but even the polarization field within the cell may be quite substantially screened, so that in fact the polarization potential in the solid state may be considerably weaker than it is in the free atom. This will have the effect of decreasing the cohesive energy.

CALLAWAY has discussed the influence of the polarization potential on the cohesive energy of metallic potassium [2]. Outside the core his potential corresponds to

$$\begin{cases} \Delta V(r) = -\dfrac{2\alpha'}{r^4} & r < r_s\,, \\ = 0 & r > r_s\,. \end{cases} \tag{6.4}$$

This is precisely the potential corresponding to the hypothesis (6.2). It cannot be reasonable because the potential is obtained by integrating the force inward

from infinity, and if the field vanishes outside the cell, then the potential (6.4) must vanish at the boundary. Thus we should add a constant potentia l $+2\alpha'/r_s^4$ to Callaway's results to conform to a model which has some possibility of physical reality. In the case of K this gives a reduction of the cohesive energy amounting to 0.0102 Ry, or 3.2 kcal. This probably represents the largest polarization potential in the solid which can reasonably be assumed, since it takes no account of screening within the cell. Nevertheless, it is very convenient for calculation, for it enables us to make computations with the model corresponding to (6.2) with only a simple additive correction to make the potential vanish at the cell boundary. As will be seen, even this model gives too little cohesion and too large an equilibrium lattice constant so that corrections for screening within the cell will make matters worse.

3. – Comparison with Callaway's results for K

It is of interest to compare results obtained by the Q.D.M. with Callaway's results using an explicit exchange and polarization potential. The agreement between the two methods for the same model is very close indeed.

For $r_s = 4.86$, before correcting the η-defects for polarization, we obtain $E_0 = -.4905$. After correction made with a cut-off radius $r_i = r_s$, the energy is raised by .0029 Ry to $-.4876$. Callaway's result obtained in the Hartree-Fock approximation, using exchange, but not polarization, is $E_0 = -.453$. The difference of .0346 Ry gives directly the lowering of the energy produced by the polarization potential, including the part inside the core. By direct calculation, using a polarization potential constructed according to a method due to Bethe, Callaway obtains an energy reduction of .0393 Ry. In quoting this result I have reduced the number actually quoted by Callaway in the ratio of the polarizability coefficients assumed outside the ion core. The value of α' used was that quoted by Ham [1]. The justification for using this value is that it is just the value required to adjust the quantum defects for the higher angular momentum states to zero, which it is easy to show they must be.

For E_2 the raw Q.D.M. calculations give 1.095 for $r_s = 4.86$. After correction of the η_1-defects we obtain a value of $E_2 = 1.149$, Callaway's value, after adjusting the polarizability as above is 1.143. Thus the agreement of the two computations is seen to be excellent.

For E_4 we obtain by the Q.D.M. a value of -0.3, which is to be compared with Callaway's value of -0.4. The agreement is again good when one considers the fact that E_4 represents a balance between nearly cancelling terms which are much larger.

The various contributions to the cohesive energy are summarized in Table XXIII.

TABLE XXIII. – *Cohesive energy of Potassium for* $r_s = 4.86$.

E_0	−.4876	
2nd order Fermi energy	.1073	
4th order Fermi energy	−.0035	
Interaction energy	.0039	(Pines, revised theory)
Polarization potential at cell boundary	.0102	
Ionization energy	.3190	
Total	.0507	

The final energy corresponds to a cohesive energy of 15.8 kcal/mol, which is to be compared with the experimental value of 21.7 kcal. The discrepancy is disappointingly large, and arises almost entirely from the various corrections and improvements to the theory that we have discussed.

These discrepancies become larger as we proceed towards the heavier elements. It is found that not only is the cohesive energy definitely too low, but also the predicted equilibrium lattice constant is too large by 10 to 15% for K, Rb and Cs. It has not been considered worth while to make any further improvements in the computation of the polarization corrections because it is evident that the present model gives about the most cohesion that can be obtained, and further improvements will all be in such a direction as to worsen the agreement with experiment both for the cohesion and for the lattice constant. For example, taking into account screening of the polarization field within the cell will raise the energy still further. It is possible that corrections to the Coulomb energy along the lines discussed in Sect. 4 of Chapter V will improve matters considerably, but these have not yet been carried out. In the case of Rb and Cs there is some doubt as to whether the expansion of the energy in powers of k^2 converges sufficiently rapidly to be reliable, although the results shown in Table XXII indicate that the power series is not too bad at $r_s = 5.6$.

In the case of Rb and Cs it is probably also necessary to take into account the van der Waals attraction of the ion cores. These will be proportional to the square of the polarizability constant α', and hence will become of increasing importance relative to the electronic polarization corrections. Furthermore a simple physical argument may be adduced to show that the van der Waals interactions are probably not screened by the valence electrons. The reason for this is simply that the frequencies involved in the electric fields which produce correlations in the electronic motion in the ion cores are much higher than the plasma frequency of the valence electrons. Thus in accordance with the ideas that will be discussed by PINES the electron gas is nearly transparent to these frequencies and will be incapable of screening the corresponding interactions. This in turn means that the van der Waals interactions may extend

to considerable distances in the crystal. Since this is an attractive interaction which varies as r_s^{-6} it wil also contribute considerably towards reducing the lattice constant. However, calculations indicate that this effect is not sufficient to compensate for all the other positive contributions to the energy discussed in the preceding paragraphs.

It is interesting to note that when calculations of the cohesive energy and lattice constant of the alkali metals are made by the Q.D.M. with none of the refinements we have been discussing in this Chapter and wit the omission of the terms originating from E_4 in the Fermi energy, the agreement with experiment for both the cohesive energy and the lattice constant is exceedingly good. The reader is referred to HAM's article [3] for details of this comparison.

REFERENCES

[1] F. S. HAM: *Solid State Physics*, vol. I, pp. 127-190, ed. by F. SEITZ and D. TURNBULL: (New York, 1955). Cfr. Table III, pp. 182-184.
[2] D. PINES: *Solid State Physics*, vol. I, pp. 368-449, ed. by F. SEITZ and D. TURNBULL: (New York, 1955). Cfr. eq. (6.8) and (6.10), p. 394.
[3] F. S. HAM: *Solid State Physics*, vol. I, pp. 368-449, ed. by F. SEITZ and D. TURNBULL: (New York, 1955). Cfr. Table V, p. 187.

VII.

Other metals.

1. – A simplified theory of cohesion.

The boundary condition for the determination of E_0 may be written in the form

$$f_0'(E_0, r_s) = 0 \,, \tag{7.1}$$

where $f_0(E_0, r)$ is the solution of the radial differential equation for $l = 0$, and the $'$ denotes differentiation with respect to r_s, explicitly. We will continue to adopt the convention introduced in Chapter III that $f_0(E_0, r)$ is normalized so that it has the value unity at r_s for all E. Equation (7.1) determines E_0 as a function of r_s and hence enables us to express f_0 as a function

of r_s only. By taking the total derivative of (7.1) with respect to r_s we then obtain the relation

$$f_0''(E_0, r_s) = \varphi_0' \frac{1}{r_s} \frac{\mathrm{d}E_0}{\mathrm{d}r_s}, \tag{7.2}$$

where φ_0' has its usual significance as defined in (3.34). We may now use the radial differential equation to find f_0'', obtaining:

$$f_0''(E_0, r_s) = V(r_s) - E_0 . \tag{7.3}$$

(Note that the $'$s always denote explicit derivatives with respect to r_s, i.e. partial derivatives keeping E_0 constant). Using (7.3) we may write (7.2) in the form:

$$\frac{\mathrm{d}E_0}{\mathrm{d}r_s} = \frac{3\gamma}{r_s}[V(r_s) - E_0], \tag{7.4}$$

where γ is the quantity defined in (3.45). This equation is completely general. It is possible to integrate it if we assume that $\gamma = 1$, as it would be for free electrons and as it nearly is for most of the alkali metals at their equilibrium lattice constant. We may also assume that $V(r_s) = -2/r_s$, i.e. that the potential is Coulombic in the outer part of the cell. Then from (7.4) we obtain:

$$E_0 = -\frac{3}{r_s} + \frac{r_0^2}{r_s^3} \tag{7.5}$$

a relation first pointed out by FRÖHLICH [1]. In this equation r_0 appears as an integration constant. We see that it is the value of r_s for which $\mathrm{d}E_0/\mathrm{d}r_s = 0$, and for which E_0, therefore, reaches its minimum value. From (7.3) we see that

$$E_{\min} = -\frac{2}{r_0} . \tag{7.6}$$

Equation (7.5) provides us with a method for finding the ground state energy as a function of r_s provided that we can determine $E_{\min}$ or r_0. A method of doing this has been suggested by KUHN [2]. We give here a slightly altered version of the Kuhn method which is more appropriate when tables of Coulomb wave functions are available, and is more accurate.

Introducing the regular and irregular Coulomb functions as in Eq. (5.9), we can express the boundary condition (7.1) in the form

$$F_0'(E_0, r_s) - \operatorname{tg} \pi\delta_0 G_0'(E_0, r_s) = 0 . \tag{7.7}$$

For the energy at the minimum we may write, taking into account (7.6)

$$\operatorname{tg} \pi \delta_0 = \frac{F_0'(-2/r_0, r_0)}{G_0'(-2/r_0, r_0)} . \tag{7.8}$$

The right hand side is a universal function of r_0 which depends only on the properties of the defined solutions of the confluent hypergeometric differential equation. Through (7.6) it may also be expressed as a universal function of E_{min}.

From the spectroscopic data on s-states we may also plot δ_0 as a function of E for each metal. When the curve of δ_0 vs. E intersects the curve of δ_0 vs. E obtained by solving (7.8) for δ_0, the minimum energy or r_0 for the particular metal is determined. Values of E_{min} for the alkali metals determined by KUHN [2] are shown in Table XXIV. Kuhn's values were obtained by

TABLE XXIV.

Element	Ionization Energy	Kuhn Min.	Brooks Min.	Ratio
Li	.396	.706	—	1.78
Na	.378	.667	.6644	1.77
K	.319	.530	—	1.66
Rb	.307	.508	—	1.65
Cs	.286	.468	.4722	1.63

the Imai WKB method as described in Chapter V, and as an indication of their accuracy we show the exact values obtained using equation (7.8). The ratio of E_{min} to the ionization energy is also tabulated, and is seen to be surprisingly constant.

The chief weakness of Equation (7.5) is that γ is assumed constant and equal to 1. By differentiating the normalization integral at $r = r_0$, it is easily shown that

$$\frac{\mathrm{d}\gamma}{\mathrm{d}r} = -\frac{3}{r}\gamma(\gamma - 1) . \tag{7.9}$$

Although this relation is only true for $r = r_0$, we may obtain a rough idea of the variation of γ with r by integrating (7.9). We obtain

$$\gamma = \frac{Ar^3}{Ar^3 - 1} , \tag{7.10}$$

where A is an integration constant. Using (7.10) in (7.4) and integrating we find

$$E_0 - \frac{r_s^3}{r_s^3 - r_1^3}\left\{-\frac{3}{r_s} + \frac{r_0^2}{r_s^3} + \frac{2\,r_1^3}{r_0\,r_s^3}\right\}, \tag{7.11}$$

where r_1 is an integration constant. This relation is found to represent the actual variation of E_0 with greater accuracy than (7.5) but has one more arbitrary constant. It may be used for interpolating if the minimum and one other energy are known.

2. – Polyvalent metals.

RAIMES [3] has suggested that polyvalent metals could be treated by considering all the valence electrons outside closed shells on an equivalent basis. The ion core field is then that of the multiply charged ion, and the solid state one-electron parameters are determined with respect to this multiply charged ion core field. The interaction of the valence electrons is taken into account entirely in the Coulomb self-energy term, which is computed in the free electron approximation. The free electron picture ought to be moderately good for such metals since the ion cores are quite small, and the valence electron wave functions are fairly constant over most of the cell.

If there are v electrons outside a rare gas configuration the radial differential equation outside the ion core takes the following form, in which we have included the potential due to the other valence electrons and also the polarization potential,

$$\frac{\mathrm{d}^2 f_l}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}f_l}{\mathrm{d}r} + \left[-\frac{1}{n^2} + \frac{2v}{r} - \frac{3v}{r_s} + \frac{vr^2}{r_s^3} + \frac{2\alpha'}{r^4} - \frac{l(l+1)}{r^2}\right] f_l = 0\,. \tag{7.12}$$

We make a change of variable

$$\varrho = vr\,, \qquad n' = vn\,, \qquad E' = \frac{E}{v^2}\,. \tag{7.13}$$

Substituting in (7.12) we find:

$$\frac{\mathrm{d}^2 f_l}{\mathrm{d}\varrho^2} + \frac{2}{\varrho}\frac{\mathrm{d}f_l}{\mathrm{d}\varrho} + \left[-\frac{1}{n'^2} + \frac{2}{\varrho} - \frac{3}{\varrho_s} + \frac{\varrho^2}{r_s^3} + \frac{2\alpha' v^2}{\varrho^4} - \frac{l(l+1)}{\varrho^2}\right] f_l = 0\,. \tag{7.14}$$

We note that this equation is exactly of the same *form* as that found for monovalent metals, energies and distances being merely rescaled. In fact (7.14) for a divalent metal may be regarded as the same equation as that for an

alkali metal, but with the units taken as: for energy, the ionization energy of the He^+ ion, and for length, the radius of the first Bohr orbit in the He^+ ion. We note that even the potential due to all the other electrons takes the form of that due to one electron and so enters in exactly the same way it does for the alkalis. The effective polarizability of the ion core is increased by a factor of v^2.

If, for the moment, we drop off the term due to the other valence electrons and the core polarizability from (7.14), we have an equation to which we can apply the quantum defect method directly. We will first apply the method described in Sect. **1** of this Chapter. The application is illustrates in Fig. 4.

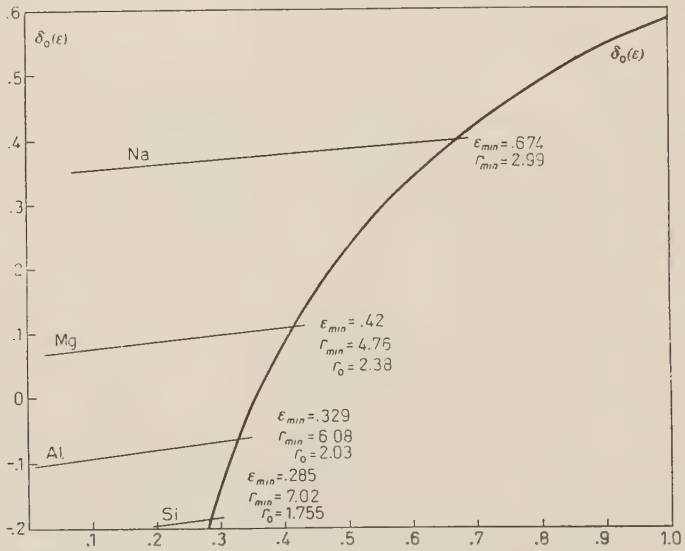


Fig. 4. – Determination of r_{min} and E_{min}.

The sloping curve represents the solution of (7.8), and the flat curves are the quantum defect extrapolations for Na, Mg, Al, and Si. It is interesting to note how the quantum defects get progressively smaller with increasing v, with the result that the extrapolation of the quantum defect becomes progressively less as v increases. However, the greater accuracy in extrapolation is probably offset by the fact that the total ground state energy is v^3 times the solution for the eigenvalue. For Si, for example, an accuracy of .001 Ry in the eigenvalue would give an accuracy of only .064 Ry (20 kcal/mol) in the final value for the cohesive energy, which is more than $\frac{1}{4}$ the observed value. It is doubtful whether the computed ground state energies can be relied on to much better than this accuracy, although this has not been estimated very carefully.

In accordance with RAIMES' suggestion that all the valence electrons may be treated as free electrons we arrive at the following general expression for the total energy of the crystal, using the approximation (7.5) for the ground state energy away from $r = r_0$

$$E = v^2\left(\frac{r_0^2}{r_s^3} - \frac{3}{r_s}\right) + \frac{1.2v^2}{r_s} + \frac{2.21v^{5/3}}{r_s^2} - \frac{.916v^{4/3}}{r_s} - vg\left(\frac{r_s}{v^{1/3}}\right), \tag{7.15}$$

where $g(r)$ is the expression for the correlation energy per electron in the free electron approximation as given for a one-electron per atom metal, e.g. by the last three terms of (2.20) or (2.21). Equation (7.15) may also be re-expressed in terms of the reduced variable ϱ as defined in (7.13):

$$E = v^3\left(\frac{\varrho_0^2}{\varrho_s^3} - \frac{3}{\varrho_s}\right) + \frac{1.2v^3}{\varrho_s} + \frac{2.21v^{11/3}}{\varrho_s^2} - \frac{.916v^{7/3}}{\varrho_s} - vg\left(\frac{\varrho_s}{v^{4/3}}\right). \tag{7.16}$$

In order, the terms of (7.15) or (7.16) have the following physical interpretation: ground state energy, Coulomb self energy, Fermi energy, exchange energy, correlation energy—all expressed in rydbergs per atom.

Typical results for the cohesive energy and lattice constant of divalent and trivalent elements are shown in Table XXV. In deriving these results equation (7.16) was used with the values of r_0 shown in Fig. 4. and the newest expression for correlation energy (2.21). The agreement between theory and experiment is seen to be quite good.

TABLE XXV.

Element	Equilibrium r_s (Bohr units)		Cohesive Energy (Ry)	
	Theory	Experiment	Theory	Experiment
Na	3.886	3.95	.0889	.084
Mg	3.437	3.30	.1159	.116
Al	3.088	2.96	.2426	.237
Si	2.75	3.15	.1872	.280
Zn	2.784	2.91	.1120	.100

It is of interest to inquire into the assumptions which enter into this approximate theory. Probably the two most questionable assumptions are that the ion core is that of the atom stripped of all its valence electrons, and that it is possible to neglect the energy discontinuities at zone boundaries and to treat the valence electrons as completely free.

The interaction between the valence electrons and the ion core is automatically taken into account in the calculation of the ground state energy.

However, there is also a contribution to the total energy which arises from the fact that the valence electrons give rise to a charge inside the ion core which partly shields it from the nucleus, thus permitting it to expand and decrease its total interaction with the nucleus. This effect decreases cohesion. The effect has been estimated by SEITZ for Li and found to be negligible, i.e. less than 0.5 kcal/mol. One might expect it to increase as we proceed to the polyvalent elements. Probably it amounts to a more or less constant percentage of the ground state eigenvalue.

With regard to the free electron approximation to the Fermi energy, we discuss this in two parts. There is, in the first place, the question of distortion of the Fermi surface near zone boundaries. This has been discussed in the last part of Chapter IV, where it was indicated that this distortion, even though large for certain k vectors, affected such a small portion of the total k-space that its influence on average quantities would be quite small.

There is, in the second place, the question of how closely the directional average of E as a function of k approximates the free electron trend. The calculations of Heine on aluminium using the orthogonalized plane wave method [4] suggest that the free electron picture is indeed a good approximation for this element except in the immediate vicinity of the zone boundaries. For the case of Mg we have attempted to compute the power series expansion of the energy as a function of k using spherical boundary conditions. For $\varrho_s = 6.60$ or $r_s = 3.30$, approximately the equilibrium sphere radius for Mg, we find

$$E_0 = -1.5020\,,$$

$$E_2 = \quad 0.9722\,,$$

$$E_4 = \quad 0.2558\,.$$

With these values the second order Fermi energy (average) is 0.314 Ry per electron, and the fourth order Fermi energy is 0.0316 Ry per electron or about 20 kcal per mole, which is almost of the order of magnitude of the total cohesion. Thus inclusion of the E_4 term in the calculation of the kinetic energy of the valence electrons destroys the good agreement between theory and experiment found in the simpler free-electron picture. The correct resolution of this dilemma is not too clear at present. It seems most likely that a calculation which took more accurate account of the effect of the potential due to the other valence electrons would tend to improve the situation, but no detailed results are available at present. It must be remarked that because of the higher average kinetic energy of the valence electrons in polyvalent metals, the cohesive energy is much more sensitive to small changes in E_2 and E_4 than in the case of the alkali metals. Thus even though the potential due to the other electrons is, relatively speaking, no more important in the polyvalent

metals than in the monovalent metals, nevertheless, the total kinetic energy is much more sensitive to these details of the potential. At the present time we must reserve judgment until further calculations are available.

3. – The monovalent noble metals.

The Q.D.M. has been applied by KAMBE [5] to the calculation of the cohesive energy of the metals Cu, Ag, and Au. The physical assumptions used were very similar to those employed by FUCHS [6] in a much earlier calculation by the Wigner-Seitz method. The distortion of the d wave functions in the solid is ignored; i.e. the d functions are considered as part of the rigid « ion core ». Departures from Coulomb potential in the vicinity of r_s were corrected for approximately by the method described in the last section of Chapter V. The empirical quantum defects were not corrected for core polarizability. This would be expected to be a larger correction in the case of Cu, Ag, and Au than for the alkalis because, as we have seen, the closed d-shells are highly polarizable. This correction, if applied, would substantially reduce the theoretically computed cohesive energy. Furthermore only the E_2 term in the expansion of the energy was computed. From the position of the atomic d levels we might expect that the E_4 term could be considerable in magnitude and positive in sign, thus further decreasing the theoretical cohesive energy. In addition to these two corrections to the one electron eigenvalues, the contribution of the repulsive energy due to ion core overlap was also neglected.

Kambe's results are summarized in Table XXVI that follows. It will be seen that even with several factors neglected which tend to reduce the theo-

TABLE XXVI.

	Theory	Experiment
Cu	59 kcal/mol	81 kcal/mol
Ag	55	68
Au	49	84

retical cohesive energy the experimental cohesion exceeds the theoretical result by a large margin in all cases. Furthermore the discrepancy becomes larger for the heavier elements having larger and more polarizable ion cores. It is also noteworthy that Ag and Au are in the wrong order theoretically.

It is difficult to see how any refinement of the type of theory described in the preceding chapters could lead to much improvement in the comparison

of theory and experiment. It seems most probable that the large additional contribution to cohesion must come from the supposedly « closed » *d* shells, probably an effect of correlation in the motions of the *d* shell electrons originating in the same general way as Van der Waals interactions, but much more complicated to describe theoretically because of the overlap of the shells. As we have already mentioned in the first chapter, the fact that multivalent metals with *d* shells behave normally as regards cohesion lends support to the idea that the unusually high cohesion of the monovalent noble metals is connected with the large overlap of the closed shells.

REFERENCES

[1] H. Fröhlich: *Proc. Roy. Soc.*, A **158**, 97 (1937).
[2] T. S. Kuhn: *Phys. Rev.*, **79**, 515 (1950).
[3] S. Raimes: *Phil. Mag.*, **43**, ser. 7, 327 (1952).
[4] V. Heine: *Proc. Roy. Soc.*, A **240**, 340, 361 (1957).
[5] K. Kambe: *Phys. Rev.*, **99**, 419 (1955).
[6] K. Fuchs: *Proc. Roy. Soc.*, A **151**, 585 (1935).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Optical Properties of Solids.

D. L. DEXTER

University of Rochester - Rochester, N.Y.

In these lectures I propose to discuss some of the important aspects of the theory of optical properties of non-metallic solids. Clearly a selection of material must be made, and I have chosen to discuss optical band-to-band transitions, properties of excitons, and the interpretation of optical properties of isolated imperfections on the basis of the « configuration co-ordinate » model. Most of what I have to say is not new, and may be well-known to some of the people attending the Summer School. My defense is that I have been instructed to give a somewhat didactic treatment. In each part some new material will be presented.

I.

Band-to-band transitions.

In order to make clear what types of transitions are involved in the first two parts, let us consider briefly the energy level structure of a non-metal. We may use the thight-binding approximation for simplicity, though the results are general. Consider N isolated identical atoms. If the i-th atomic energy level has degeneracy g_i, there is Ng_i-fold degeneracy in the corresponding energy level of the system, because of the equivalence of the N atoms. If the atoms coalesce to form a solid, most of the degeneracy is removed by the interactions of the atoms, and each level is broadened into a band. The levels corresponding to the outermost valence electrons make up the valence band (or bands), which is (are) of course completely filled if the system is a non-metal. Each valence electron occupies a state in the one-electron approximation of the form $\psi_{\boldsymbol{k}_i} = \exp[i\boldsymbol{k}_i \cdot \boldsymbol{r}]\chi_{\boldsymbol{k}_i}(\boldsymbol{r})$, where $\chi_{\boldsymbol{k}_i}$ has the periodicity of the lattice.

If one additional electron is introduced into the crystal when it is in its ground state, the electron experiences a periodic potential, and has available

a band of conducting levels which it may occupy. These states are again describable by a wave function of the form $\psi_{k_c}(\boldsymbol{r}) = \exp[i\boldsymbol{k}_c \cdot \boldsymbol{r}]\chi'_{k_c}(\boldsymbol{r})$. It is these states which would normally be computed and called the conduction band. It is important to recall that these are states for an additional electron.

The excited but non-ionized atomic levels likewise have most of their degeneracy lifted, and bands of excited, non-conducting states *of the system* result.

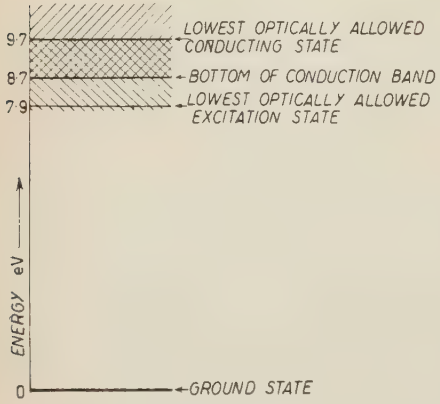


Fig. 1. – Schematic energy level diagram of an insulating crystal in which the lowest excited states are not conducting states. These are states of the entire system, not one-electron states.

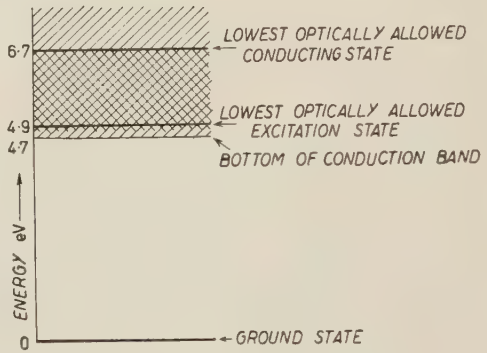


Fig. 2. – Schematic energy level diagram of an insulating crystal in which the conducting states have overlapped the « exciton » states at the existing inter-atomic separation.

These states are the exciton, or excitation states, and are non-existent in the Bloch description. In addition there will be conducting states, similar to the ionization continuum of the atomic levels. Fig. 1 and 2 show these states schematically.

For large separations of the atoms these conducting levels will in general be of higher energy than the excitation states. At the actual atomic separation that exists in the solid they may or may not begin at a higher energy. If they are of higher energy, the onset of optical absorption will be to the excitation states, and photoconductivity will not result. We shall discuss this case later on. If the conducting states begin at lower energy, it is natural to assume that these states are practically identical to the states computed for the additional electron, as described above. This is the basis for the usual theory of band-to-band transitions in solids. An essential part of this procedure is the neglect of the effect of the removal of a valence electron upon the energies and wave functions of the excited one-electron states.

It is possible that in all cases the lowest energy excited states are of the non-conducting exciton type; although no general proof of this theorem has

been given, it can be derived under special conditions. In some of the semiconductors the optical absorption can be reasonably well understood on the basis of band-to-band transitions alone, with no consideration of exciton transitions. Hence the remainder of this section is intended to apply to systems in which either the exciton states occur at higher energy than the onset of band-to-band transitions, or there is negligible oscillator strength associated with the excitation states.

1. – Absorption in a static crystal.

1·1. *Basic equations.* – Let us proceed to a discussion of the shape of the absorption coefficient to be expected in a perfect crystal, for band-to-band transitions. The absorption coefficient μ is defined by the relation

$$\mathrm{d}I = -I\mu\,\mathrm{d}x\,, \tag{I-1}$$

where I, the light intensity, is the energy flux, or the energy density times the velocity of energy flow. Thus the absorption coefficient is the energy removed per unit time per unit volume from a beam of unit intensity. Let us define $\boldsymbol{\epsilon}$ as the electric field within the medium of dielectric constant $\varkappa_0$, so that $\varkappa_0\varepsilon^2/(4\pi)$ is the energy density, v the energy velocity, w the transition probability per unit time per unit volume, and $E = \hbar\omega$ the energy of the transition. We obtain

$$\mu(E) = \frac{4\pi E}{\varkappa_0\varepsilon^2 v}\, w(E)\,. \tag{I-2}$$

We treat a crystal of volume V and N atoms, and must obtain, using time-dependent perturbation theory, the transition probability w. In the absence of selection rules the absorption of a quantum of light can cause an electron in the valence band in any state $\boldsymbol{k}_v$, of energy $\varepsilon(\boldsymbol{k}_v)$, to be raised to any state $\boldsymbol{k}_c$, of energy $u(\boldsymbol{k}_c)$, in the conduction band. The probability that this process has occurred by time t is denoted by $|a_{\boldsymbol{k}_c,\boldsymbol{k}_v}(t)|^2$, where $a_{\boldsymbol{k}_c,\boldsymbol{k}_v}(t)$ is the time-dependent amplitude of the excited state characterized by a hole in the valence band with propagation vector $\boldsymbol{k}_v$ and an electron in the conduction band described by $\boldsymbol{k}_c$. Summing over all pairs of states, we find the probability per unit volume per unit time that the system has been raised to a conducting state by the absorption of a photon

$$w = 2(Vt)^{-1} \sum_{\boldsymbol{k}_c} \sum_{\boldsymbol{k}_v} |\langle \boldsymbol{k}_c,\ \boldsymbol{k}_v | H_R | 0 \rangle|^2 \left| \frac{4\sin^2 (t/2\hbar)(\hbar\omega_{\boldsymbol{k}_c,\boldsymbol{k}_v})}{(\hbar\omega_{\boldsymbol{k}_c,\boldsymbol{k}_v})^2} \right| . \tag{I-3}$$

The factor of 2 arises because of spin, and the summations over $\boldsymbol{k}_v$, and $\boldsymbol{k}_c$ include all occupied and unoccupied space states. The function in brackets represents the familiar selection rule on the conservation of energy, giving an appreciable contribution only when

$$\hbar\omega_{\boldsymbol{k}_c,\boldsymbol{k}_v} \equiv u(\boldsymbol{k}_c) - \varepsilon(\boldsymbol{k}_v) - \hbar\omega \tag{I-4}$$

is approximately zero. The Hamiltonian describing the interaction of the electrons with the radiation field is given by the sum over all electrons,

$$H_R = \frac{ie\hbar A_0}{mc} \sum_i \boldsymbol{a}\cdot\boldsymbol{\nabla}_i \exp\left[i\boldsymbol{\eta}\cdot\boldsymbol{r}_i\right], \tag{I-5}$$

where $\hbar\boldsymbol{\eta}$ is the momentum and $\hbar\omega$ the energy of the photon associated with the vector potential

$$\boldsymbol{A} = A_0\boldsymbol{a}\exp\left[i(\boldsymbol{\eta}\cdot\boldsymbol{r} - \omega t)\right] + \text{C. C.}, \tag{I-6}$$

with a polarization described by the unit vector $\boldsymbol{a}$.

Computing the matrix element, we find

$$\langle \boldsymbol{k}_c, \boldsymbol{k}_v | H_R | 0\rangle = \frac{ie\hbar}{mcV} A_0 M \sum_{\boldsymbol{R}_J} \exp\left[i(\boldsymbol{k}_v - \boldsymbol{k}_c + \boldsymbol{\eta})\cdot\boldsymbol{R}_J\right], \tag{I-7}$$

where M is similar to an atomic matrix element, being an integral over a unit cell,

$$M = \int \left\{\chi'^{*}_{\boldsymbol{k}_c}[\boldsymbol{a}\cdot\boldsymbol{\nabla} + i\boldsymbol{a}\cdot\boldsymbol{k}\,]\chi_{\boldsymbol{k}_v}\right\}\exp\left[i(\boldsymbol{k}_v - \boldsymbol{k}_c + \boldsymbol{\eta})\cdot\boldsymbol{r}\right]\mathrm{d}\tau\,. \tag{I-8}$$

The sum in Eq. (I-7) is very close to zero unless

$$\boldsymbol{k}_v - \boldsymbol{k}_c + \boldsymbol{\eta} = 0\,, \tag{I-9}$$

leading to the selection rule on conservation of momentum. $\boldsymbol{\eta}$ is small for all cases of interest here, being several orders of magnitude smaller than most values of $\boldsymbol{k}_c$, $\boldsymbol{k}_v$. We may accordingly set $\boldsymbol{\eta} = 0$ for our present purposes, leading to the approximate statement that a transition is forbidden when the electron changes its momentum. Thus the sum in Eq. (I-7) may be evaluated as $N\delta_{\boldsymbol{k}_v - \boldsymbol{k}_c, 0}$ and the sum over $\boldsymbol{k}_c$ in Eq. (I-3) may be carried out by means of the Kronecker delta function. Changing the other summation to an integral, we obtain

$$\mu(E) = \frac{1}{t}\,\frac{2e^2\hbar^3}{\pi^2\varkappa_0 m^2 v}\left(\frac{N}{V}\right)^2 \frac{1}{\omega}\sum \int M^2 \frac{\sin^2(t/2\hbar)(\hbar\omega_{\boldsymbol{k}_c,\boldsymbol{k}_v})}{(\hbar\omega_{\boldsymbol{k}_v,\boldsymbol{k}_v})^2}\,\mathrm{d}^3\boldsymbol{k}\,. \tag{I-10}$$

The summation in Eq. (I-10) is a sum over all the valence bands which may be of interest in the energy range covered.

To carry the calculation further, the problem must be specialized in terms of the energy band structure and the Bloch functions χ'_{k_c} and χ_{k_v}.

1·2. *Results for specific band structures.* – The simplest model we treat is the case of spherical energy surfaces, with a minimum in the conduction band and a maximum in the valence band for zero momentum, at which the energy separation is U_0. If m_c and m_v are the effective masses, with $\gamma = m_c/m_v$ their ratio, we may determine the integrand of Eq. (I-10) and evaluate the integral. We must first inspect the matrix element M in Eq. (I-8), however. For small values of $\boldsymbol{k}_v$ in Eq. (I-8) we would expect the $\boldsymbol{a}\cdot\nabla$ term to dominate over the $\boldsymbol{a}\cdot\boldsymbol{k}_v$ term, so that M would be almost independent of $\boldsymbol{k}_v$ for $\boldsymbol{k}_v$ small. This will be the case if the symmetries of χ_{k_v} and χ'_{k_c} are such that the transition is an allowed one. For example, if χ_{k_v} is like an atomic s-state and χ'_{k_c} like a p-state, the matrix element of $\boldsymbol{a}\cdot\nabla$ would be non-zero, and would be the largest contribution to M. If the symmetries were both s-like, however, the $\boldsymbol{a}\cdot\nabla$ matrix element would be zero, and the $\boldsymbol{a}\cdot\boldsymbol{k}_v$ contribution would dominate. We will refer to these cases as symmetry-allowed and symmetry-forbidden transitions. These selection rules are in addition to the energy and momentum selection rules mentioned previously.

On the assumption that the transition is symmetry-allowed, we may remove M^2 from the integral of Eq. (I-10). The evaluation is straightforward, and results in the absorption coefficient,

$$\mu(E) = \frac{2^{\frac{3}{2}}}{\varkappa_0 v}\left(\frac{N}{V}\right)^2 \frac{e^2 m_c^{\frac{3}{2}} M^2}{(1+\gamma)^{\frac{3}{2}}} \frac{(E-U_0)^{\frac{1}{2}}}{E}, \tag{I-11}$$

for $E - U_0$ small. That is, near the onset of absorption at U_0, a square root dependence on $E - U_0$ should occur. This dependence comes essentially from the density of states in energy space. For larger values of $E - U_0$ the dependence would be more complex, depending on the details of the band structure. There would in general be a maximum in the μ vs. E curve at $E - U_0$ of the order of a very few eV, where μ reaches a value of the order of 10^6 cm^{-1}, and a decrease at higher energies.

If the transition were symmetry-forbidden, on the other hand, we see that M^2 would be proportional to $\boldsymbol{k}_v^2$, which is also proportional to $E - U_0$. Thus we would expect a $\frac{3}{2}$ power dependence of μ on $E - U_0$.

It is important to note that with this energy model, the lowest possible energy transition is also an allowed transition according to the momentum selection rule. Thus imperfections which would destroy momentum as a good quantum number would have relatively little influence on the absorption band shape.

Other assumed band structures give similar results. For example, if the valence band and conduction band both have minima at $\boldsymbol{k} = 0$, with effective masses $m_c < m_v$, we obtain the same results as in Eq. (I-11), but with $1+\gamma$ replaced by $1-\gamma$. Thus if the effective masses are very closely the same, all momentum-allowed transition would have approximately the same energy, and a very sharp absorption edge would be observed. In this case, however, if the momentum selection rule were broken down, transitions corresponding to $\boldsymbol{k}_v \neq 0$, $\boldsymbol{k}_c = 0$ could occur at lower energy than U_0, and a low-energy tail should occur.

Still other simple band structures may be readily treated, but in each case one obtains the one-half (or three-halves) power dependence previously noted, with a proportionality constant and a distance from U_0 to the maximum of the curve depending on the details of the band structure. There does not seem to be any known case in which this theory is sufficient to explain the observed absorption. In all cases except those for which the maximum energy for the valence band occurs at the same value of $\boldsymbol{k}$ as the minimum for the conduction band, momentum-destroying transitions can take place to extend the long wavelength tail of the absorption band [1].

1·3. *Temperature and pressure dependence.* – In the zeroth approximation, in which we do not consider lattice vibrations, the only way in which the absorption coefficient changes with temperature is through the dependence of the band gap U_0 upon temperature. (This dependence is of course connected with lattice vibrations, but such a connection might be ignored in a phenomenological treatment.) Thus we might attempt to interpret the variation in absorption with temperature in terms of a deformation potential theory,

$$U_0(T) = U_0(0) + E_{Ig}\Delta(T)\,, \tag{I-12}$$

where $\Delta(T)$, the average dilatation, may be related to the lattice constant in an obvious fashion. $U_0(0)$ is the band gap at absolute zero, and E_{Ig} is the variation in band gap per unit dilatation. E_{Ig} may be obtained from pressure measurements independently. There seems to be no known case where such a treatment of the temperature dependence is adequate [2], although the pressure dependence seems reasonable.

1·4. *Effects of static imperfections.* – Static imperfections can modify the absorption in two ways, either by breaking down momentum selection rules or by creating new energy states not existent in the perfect infinite lattice. The surface of the crystal is of course an unavoidable imperfection, particularly important where absorption is strong and takes place near the surface. With even a perfect surface, apparent absorption coefficients of the order of 10^1 cm^{-1}

may be found in a region almost transparent in the large single crystal. In actual thin films, where the surface is far from perfect, still larger effects are frequently observed [3].

Effects of dislocations on breaking down momentum selection rules have also been considered. It appears that in crystals of reasonable perfection the effects of lattice vibrations, to be discussed later, would generally dominate dislocations in this respect. New states created by the presence and motion of dislocations would be expected to be important, however, particularly in ionic crystals [4]. Other types of imperfection would have an influence also, but usually more closely associated with localized effects than with band-to-band transitions.

2. – Effects of lattice vibrations.

It has become increasingly apparent in recent years that observed absorption spectra are intimately connected with the interaction between electrons and lattice vibrations. Several distinct effects have been considered, but no complete theory has yet evolved.

2·1. *Thermal lattice expansion.* – One relatively trivial effect has already been mentioned, namely, the usual shift toward higher energy of the « edge » of the absorption band with decreasing temperature, associated with the decreasing lattice constant. There can be little doubt that such an effect exists, but it surely is by no means the only one [2].

2·2. *Thermal broadening of levels.* – A second effect is the broadening of each one-electron level, according to the Uncertainty Principle, by the collisions of the electron with lattice vibrations. In polar crystals, where the interaction is large, the scattering time, hence the lifetime of the states, is small, and this effect has been shown to be quite significant [5].

2·3. *Thermal shifting of levels.* – A vibrating lattice not only scatters electrons, but also interacts with them, the interaction energy depending on the state of the electron. Thus when an electron changes its state in an optical transition, the energy of the transition is modified by the difference in the interaction energies associated wtih the two states. This effect has been shown to be comparable with that of thermal expansion in non-metals [2].

2·4. *Violation of momentum selection rules.* – Perhaps the most useful theory of temperature effects is that of the « indirect transitions ». As discussed briefly above, if the minimum of the conduction band occurs at a different

value of momentum from that of the maximum of the valence band, a breakdown of the momentum selection rule allows low-energy transitions not otherwise possible. These « non-vertical » transitions have been called indirect transitions. A treatment has been given of the effect of lattice vibrations on these transitions, in which it is shown that absorption or emission of a phonon of suitable momentum enables these non-vertical transitions to occur. This theory [6], with slight modifications [7], has been applied with at least qualitative success in some of the semiconductors. Although this topic deserves considerable attention, I will not go into detail here, as Professor FAN has covered it in his lectures.

2 5. *Creation of new states.* – The effects mentioned above are all computed using the electronic states which would exist in the periodic non-vibrating lattice, emphasis being variously placed on the shifting or broadening of the levels, or on the breakdown of selection rules. Thus each effect is described in terms of a perturbation treatment.

There exists a further and complex effect not previously discussed, to the best of my knowledge. This effect is concerned with the new electronic states which must be taken into account when we consider the non-periodicity of the lattice potential in a vibrating crystal. Instead of a sharp cut-off in the density of states for a periodic potential

$$\begin{aligned} \varrho(E) &= \propto [E - U_0(T)]^{\frac{1}{2}} && \text{for } E \geqslant U_0(T). \\ &= 0 && \text{for } E < U_0(T), \end{aligned}$$

where $U_0(T)$ is the minimum band gap in the crystal with a given lattice constant $\lambda(T)$, there must occur in the vibrating crystal a non-zero density of states at energies appreciably less than $U_0(T)$. (A similar conclusion is reached in a consideration of the density of states associated with an impurity band, when the impurities are located randomly, rather than on a super lattice [8].) It might at first be thought that the effect would be small, but if we are concerned with the behaviour near the band edge, say at $\mu \sim 1$ cm^{-1}, we must recognize that we are considering a value of μ of the order 10^{-6} of the maximum value, so that even a small effect may be important in this range.

Another point of view is that for energies appreciably less than $U_0(T)$, perturbation theory cannot tell us the density of states. Unfortunately the exact solution of the problem is beset by impressive mathematical difficulties. The one-dimensional problem of the density of states in a not-quite-periodic potential, although distinctly non-trivial, has recently been solved [9]. There is no apparent generalization to three dimensions.

In order to demonstrate that the problem is not only difficult but im-

portant, let us mention recent observations on the absorption curves of the silver halides [10] and of cadmium sulfide [11]. Urbach originally made the observation that in the silver halides the absorption coefficient μ rose exponentially with increasing energy for several decades in the vicinity of $\mu = 1\ \mathrm{cm}^{-1}$. He further noted that the coefficient was proportional to $1/kT$. That is, in these materials $\mu = K \exp[E/ckT]$, where c is a dimensionless quantity close to unity. In CdS the relation is obeyed even more strikingly, being observed over a longer range of absorption coefficient, and the constant K varying by about a factor of 4, in polarized light, depending on the orientation of the electric vector with respect to the crystallographic c-axis of the crystal. Such a result as « Urbach's Rule » is obviously not interpretable on the basis of the theories previously discussed, and suggests the existence of an exponential tail in the density of states toward lower energy measured from the band gap $U_0(T)$.

Let us consider how such an exponential tail might arise. We treat for simplicity the case of symmetry-allowed, « direct » transitions so that the existence of imperfections cannot produce a low-energy tail by destroying momentum selection rules. In this case, as we have seen, μ is approximately proportional to the density of states, which in the static lattice is again proportional to $[E - U_0(T)]^{\frac{1}{2}}$. That is, according to Eq. (I-11), $\mu = \mu_0[E - U_0(T)]^{\frac{1}{2}}$, and we will neglect the slight energy dependence in μ_0. However, there exist local variations in the band gap because of lattice expansions and contractions associated with lattice vibration. The mean square dilatation associated with thermal effects is equal to $\langle \Delta^2 \rangle_{\mathrm{Av}} = kT/C_{ii}$ at temperatures where equipartition of energy obtains, where C_{ii} is an appropriate elastic constant. Now if the spatial extent of the local dilatation were large with respect to the wavelength of the electrons undergoing the optical transition, we could use deformation potential theory to relate the local change in band gap $W(\boldsymbol{r})$ to $\Delta(\boldsymbol{r})$ through the relation $W(\boldsymbol{r}) = E_{1G}\Delta(\boldsymbol{r})$, and thus obtain $\langle W^2 \rangle_{\mathrm{Av}} = E_{1G}\langle \Delta^2 \rangle_{\mathrm{Av}} = E_{1G}kT/C_{ii}$. (This assumption regarding the spatial extent of the local dilatations is clearly incorrect, and will be returned to later.) Further, at reasonably high temperatures the atoms may be considered to oscillate independently and the relative probability of a given change in the band gap, W, where the mean square value is $\langle W^2 \rangle = E_{1G}^2 kT/C_{ii}$, is given approximately by

$$P(W) = \left[\frac{C_{ii}}{2\pi E_{1G}^2 kT}\right]^{\frac{1}{2}} \exp\left[-W^2 \frac{C_{ii}}{2E_{1G}^2 kT}\right]. \tag{I-13}$$

Thus, taking into account the local variations in band gap, we find the absorption coefficient to be given by

$$\mu(E) = \mu_0 \left[\frac{C_{ii}}{2\pi E_{1G}^2 kT}\right]^{\frac{1}{2}} \int_{-\infty}^{\infty} \mathrm{d}W [E - (U_0 + W)]^{\frac{1}{2}} \exp\left[-W^2 \frac{C_{ii}}{2E_{1G}^2 kT}\right]. \tag{I-14}$$

Thus, for energies E appreciably less than the edge in the static crystal U_0, we find

$$\text{(I-15)} \quad \mu(E) = \mu_0 \frac{E_{IG}^2 kT}{2^{\frac{3}{2}} C_{ii} (U_0 - E)^{\frac{3}{2}}} \exp\left[- \frac{C_{ii}}{2E_{IG}^2 kT} (U_0 - E)^2 \right] \cdot \left[1 - \frac{15}{16} \left(\frac{2E_{IG}^2 kT}{C_{ii}(U_0 - E)^2} \right) + \frac{15 \cdot 63}{512} \left(\frac{2E_{IG}^2 kT}{C_{ii}(U_0 - E)^2} \right)^2 + \dots \right].$$

The logarithmic derivative of the absorption coefficient, which according to Urbach's Rule should be $(ckT)^{-1}$, where $c \cong 1$, we find to be

$$\text{(I-16)} \quad \frac{\mathrm{d} \ln \mu}{\mathrm{d}E} = \left\{ \frac{C_{ii}(U_0 - E)}{E_{IG}^2} \left[1 + \frac{3}{4} \left(\frac{2E_{IG}^2 kT}{C_{ii}(U_0 - E)^2} \right) \right] + \dots \right\} \frac{1}{kT}.$$

This expression should be modified in a region where the expansion parameter $2E_{IG}^2 kT / C_{ii}(U_0 - E)^2$ becomes comparable with or greater than unity. For $E - U_0$ positive the expression (I-14) becomes similar to the relation for the static lattice $\mu = \mu_0 (E - U_0)^{\frac{1}{2}}$.

The quantity in curly brackets in Eq. (I-16), which we refer to as A, is a dimensionless quantity that must be compared with unity. Over a small energy range well removed from the static band edge, where $U_0 - E$ is of the order tenths of an eV, A is roughly constant for small variations in E. For typical values of elastic constants and deformation potential constants, the coefficient A is found to be appreciably less than unity. This is not an unexpected result. That is, since the spatial extent of the local dilatation is *not* large with respect to an electron's wavelength, the electron will naturally tend to cancel out the effects of positive and negative dilatations. We would thus expect to have to use smaller values of deformation potential constants than those determined in a uniform dilatation experiment. If E_{IG} were reduced by a factor of order 5, reasonable agreement would be obtained.

Specifying values of E_{IG} and C_{ii} such that A in Eq. (I-16) has the value unity for $U_0 - E$ equal to 0.5 eV, we integrate Eq. (I-14) and obtain the solid curve in Fig. 3. The dashed

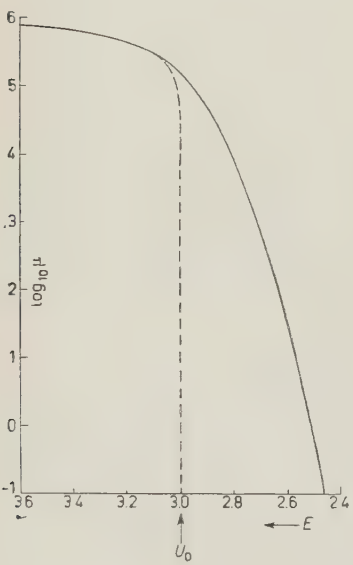


Fig. 3. – Calculated absorption spectrum for band-to-band transitions in a non-periodic potential, shown with the solid line, and in a periodic potential, with the same band gap, with the dashed line. The apparent band gap moves by about half an electron volt.

curve would be obtained with the static lattice. We have selected the reasonable value $10^6\,\mathrm{cm}^{-1}\,\mathrm{eV}^{-\frac{1}{2}}$ for μ_0, and have arbitrarily chosen a value 3.0 eV for U_0. The temperature is taken to be room temperature. It will be noted that for μ less than about $10^3\,\mathrm{cm}^{-1}$, the absorption coefficient is reasonably close to a straight line of slope $1/kT$. It is important to note

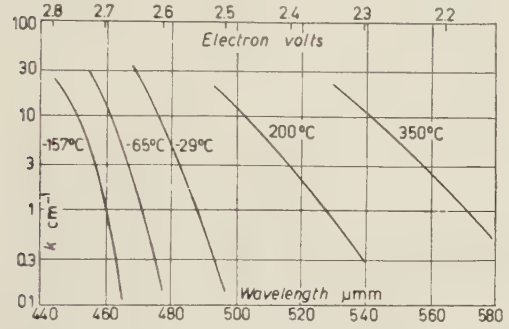


Fig. 4. – Observed absorption coefficient for a single crystal of AgBr as a function of temperature. (After Moser and Urbach, reference [10].) Note the $1/T$ dependence in the slope at high temperature.

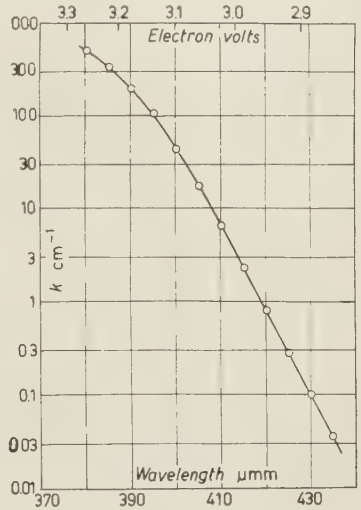


Fig. 5. – Observed absorption coefficient for a single crystal of AgCl. (After Moser and Urbach, reference [10].)

that appreciable absorption extends to 0.5 eV lower energy than the static band gap. This result points out the great difficulty in assigning a definite number to the « optical band gap. » Experimental curves for AgBr[10], AgCl[10] and CdS[11] are shown in Fig. 4, 5 and 6, and qualitative similarity is noted. The observed dichroism for CdS may reasonably be explained if the transition matrix element parallel to the C-axis is a factor of two smaller than that perpendicular to the C-axis.

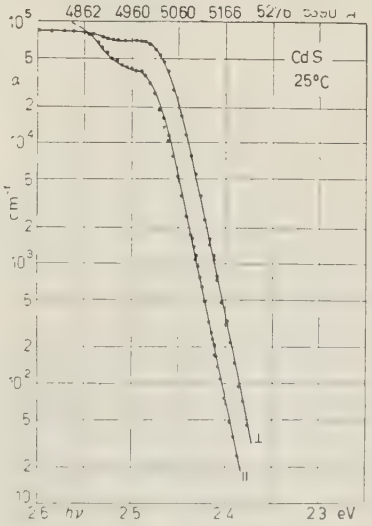


Fig. 6. – Optical absorption observed in CdS. (After Dutton, reference [11].) Note the pronounced dichroism and the parallel straight lines for several decades on the logarithmic plot. Single crystals of thickness varying from about one mm to less than one μm were used to cover the entire range.

The above arguments cannot be said to constitute a theory of the observed effect, but they do appear to furnish some basis for its interpretation. What experience there is indicates that random fluctuations in potential give rise to tails on the distribution of states which are somewhat similar to a Poisson distribution or Gaussian.

In general, the more rapidly varying is μ with $(E - U_0)$ in the static lattice, the smaller is the effect discussed here. If μ varies as $\mu_0(n)(E - U_0)^n$, increasing n by unity decreases the magnitude of the tail by about one order of magnitude. The equations corresponding to Eqs. (I-15) and (I-16) are, for the general case,

$$\text{(I-15')} \qquad \mu = \mu_0(n)\left(\frac{2E_{IG}^2 kT}{C_{ii}}\right)^{n/2}\frac{\Gamma(n+1)}{\pi^{\frac{1}{2}}2^{n+1}}\left[\frac{2E_{IG}^2 kT}{C_{ii}(U_0-E)^2}\right]^{(n+3)/2}\cdot$$
$$\cdot\exp\left[-(U_0-E)^2\frac{C_{ii}}{2E_{IG}^2 kT}\right]\left\{1-\frac{(n+1)(n+2)}{4}\left[\frac{2E_{IG}^2 kT}{C_{ii}(U_0-E)^2}\right]+\cdots\right\},$$

$$\text{(I-16')} \qquad \frac{\mathrm{d}\ln\mu}{\mathrm{d}E} = \left\{\frac{C_{ii}(U_0-E)}{E_{IG}^2}\left[1+\frac{n+1}{2}\left(\frac{2E_{IG}^2 kT}{C_{ii}(U_0-E)^2}\right)\right]+\cdots\right\}\frac{1}{kT}.$$

From Eq. (I-16') we see that the slope of the curve of $\log\mu$ versus energy is roughly independent of n, and from Eq. (I-15') that the magnitude of μ at a given energy $E < U_0$ decreases with increasing n. Thus the low energy tail is substantially reduced for symmetry-forbidden and for « indirect » transitions, at a given energy interval from the edge.

The existence of the new states associated with the vibrating lattice likewise have an influence on the determination of the band gap by the measurement of intrinsic conductivity. The influence is appreciably less, however. This effect is not included in the scope of this lecture, and will not be discussed here.

II.

Excitons.

A great deal of attention has been devoted to the subject of excitons, particularly theoretical attention. A remarkable degree of ignorance remains, however, about the actual properties associated with these states in existing crystals.

We may refer to the beginning of the first lecture for the origin of these states. To summarize, these excitation states arise in the tight binding approximation from the excited, non-ionized atomic states of the constituent atoms or molecules of the crystal. In another description the states arise from the interaction between a hole in the valence band and an electron in the conduction band. If the potential is periodic, these states must be charac-

terized by a momentum $\hbar\boldsymbol{k}$ and a propagation vector $\boldsymbol{k}$. It is important to note that these states are states of the whole system, not one-electron states, and do not exist in the Bloch scheme.

In the Heitler-London description a given state is represented by a linear combination of excited atomic states based on the many unit cells of the crystal, the coefficients differing only in phase, being given in terms of the propagation vector. Thus we may think of the electronic excitation energy as flowing from one atom to the next in a regular, coherent way. In the alternative description we imagine an electron and hole rotating about each other, similar to a positronium atom, the center of mass moving through the crystal with a definite momentum.

Most theoretical work has been based on one or the other of these schemes, in which the periodicity of the lattice is an essential assumption. Many difficulties arise when this assumption is relaxed.

There can be little doubt of the existence of excited non-ionized states of insulating crystals, when the constituent atoms or molecules interact weakly, and when their ionization potential is considerably greater than the excitation energy to their first excited states. Some of the real questions which do arise refer to the best method of description of these states, to the degree of coherence of the states, and to the extent to which energy may actually be propagated. These matters are closely connected with the amount of interaction with lattice vibrations and other imperfections, or alternatively, with the deviations from perfect periodicity of the lattice potential.

Perhaps it should be restated that, according to our present knowledge, the lowest excitation states may be of higher energy than the lowest conducting states, even in the perfect crystal. It may be that this is the case in the silver halides, but probably not in the alkali halides and many organic crystals. The occurrence of conducting states as the lowest energy excited

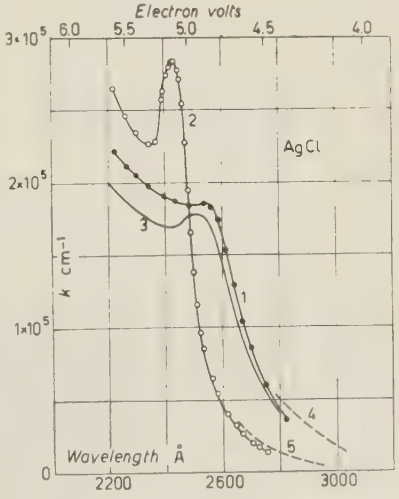


Fig. 7. – The absorption spectrum of AgCl at room temperature and liquid nitrogen temperature. Curve 1 is the absorption curve at 26 °C, and curve 2 is at — 184 °C. The measurements were taken on a thin (210 μmm) evaporated film. Curve 3 shows the absorption curve measured by Milliman at room temperature on fused films. Curves 4 and 5 are data measured by Kayser on an evaporated film of 510 μmm thickness at 27 °C and at — 183 °C respectively. (After Tutihasi, reference [12].)

states does not imply that the excitation states are non-existent, nor that their existence cannot manifest itself in optical absorption or in other ways. In such a case there may be a concentration of oscillator strength in a given energy region superimposed on the absorption arising from band-to-band transitions. However, in such a case one would expect auto-ionization of the exciton so produced in a time of the order 10^{-12} s, resulting in the creation of a free electron and a free hole. The peaks in absorption of the silver halides [12], at energies about 2 eV from the onset of strong absorption, very likely correspond to transitions to excitation states, but no diminution in photoconductivity should be observed other than that associated with surface phenomena. Fig. 7 and 8 show these peaks in two different kinds of plots.

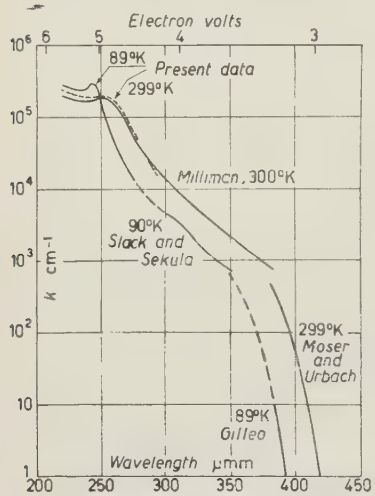


Fig. 8. - Optical absorption in AgCl over a large range of absorption coefficient at room temperature and at liquid nitrogen temperature. (After Tutihasi and others, reference [12].) Evaporated films and single crystals were used in the different energy ranges. The strong absorption begins at energies about 2 eV lower than the exciton peaks.

Just as in band-to-band transitions, here also there are energy and momentum selection rules, and, of course, selection rules arising from the symmetry of the atomic states involved in the optical transition. To demonstrate these selection rules we treat an idealized case with the tight binding approximation.

1. – Wave functions and energy levels.

Let us consider N isolated, identical atoms or molecules, each having a Hamiltonian H_I and states described by wavefunctions ψ^n and energies ε_n. The Hamiltonian for the system is then $H_0 = \sum_I H_I$, with wave functions given by a product of the various ψ^n. The ground state of the system is described by

$$\Psi^0 = \psi_1^0 \psi_2^0 \dots \psi_N^0 , \tag{II-1}$$

where the subscript labels the atom and the superscript the atomic energy

level. The lowest excited states of the system are of the form

$$\Psi_L^n = \psi_1^0 \psi_2^0 \cdots \psi_L^n \cdots \psi_N^0 , \tag{II-2}$$

in which the L-th atom is in the n-th excited state, all other atoms being in the ground state. In order to satisfy the Pauli exclusion principle, we would have to anti-symmetrize Eqs. (II-1) and (II-2). Other states exist, of course, in which two or more atoms are excited. The energy of the ground state is equal to $E_0 - N\varepsilon_0$, and is in general non-degenerate; the energy corresponding to (II-2) is $E_n = (N-1)\varepsilon_0 + \varepsilon_n$ and is at least N-fold degenerate, since the atoms are identical. If the atoms or molecules are brought together into a crystal, so that they interact, the system's Hamiltonian is given by

$$H = H_0 + \sum_I \sum_J{}' H_{IJ} , \tag{II-3}$$

where H_{IJ} represents the interaction between the I-th and J-th atoms. Functions of the form (II-2) are no longer eigenfunctions of (II-3), because of the perturbation resulting from the interactions of the atoms. Instead, most of the degeneracy is removed, and the wave functions of the system are of the form $\Phi_0 = \Psi^0$ for the ground state, and

$$\Phi_n(\boldsymbol{k}) = N^{-\frac{1}{2}} \sum_L \exp\left[i\boldsymbol{k}\cdot\boldsymbol{R}_L\right]\Psi_L^n \tag{II-4}$$

for the excited states. Again the Pauli principle should be taken into account, and in addition other correction terms should be included for most purposes. For a qualitative discussion these wave functions are sufficient [13]. Eq. (II-4) indicates that the excitation energy is being propagated from one atom to the next with a propagation vector $\boldsymbol{k}$ (and a momentum $\hbar\boldsymbol{k}$). FRENKEL introduced the term « exciton » to describe pictorially the particle-like aspect of this quantum of electronic excitation energy in a momentum state [14]. It seems to me unfortunate that the term exciton is now applied to *any* quantum of electronic excitation energy, regardless of whether or not the system is in a momentum state or the potential is periodic. Important physical problems hinge on this distinction, and the whole method of description of energy transfer, to give just one example, depends upon it. It is regrettable that Frenkel did not coin two additional words, one generic and one corresponding to a description in configuration space, to handle this difficulty. It is likely that the present usage is too firmly entrenched to be changed now, and I shall follow the common usage in the following. I hope that it will be recalled that totally different concepts are invoked in various uses of the same term.

It is the interaction among atoms or molecules which is responsible for the migration or propagation of the excitation energy, and this same interaction determines the energy spectrum. For many years it was believed that

« exchange », that is, electron exchange arising in connection with antisymmetrized wave functions, was responsible for this interaction, and was a necessary condition for transfer to occur. Following a suggestion by Seitz-Heller and Marcus showed that the electromagnetic « near-zone » interaction among excited atoms could furnish a strong mechanism for propagation [15]. This mechanism, which is essentially classical in nature, as contrasted with « exchange », is the same as the interaction between oscillators which leads to the resonant transfer of energy between neighboring identical antennae. That is, it is the dipole-dipole interaction (when the transition is an « allowed » one) which arises from the overlapping of the dipole fields of the excited molecules, rather than the actual overlapping of the electronic wave functions, which is responsible for the propagation of excitation in many cases of interest, where the excited state is reached by an optical transition. A crude way of visualizing the process of the propagation of energy is to consider that a given excited atom jumps to the ground state, passing its energy along to its neighbor in a perfectly regular, coherent way, through the interaction of their electromagnetic dipole fields. In the following we shall treat the dipole-dipole (d-d) coupling as the dominant effect, though exchange may also be important. It is important to note, as stressed in reference [15], that the d-d interaction is a long range one, and it is not sufficient to include only near-neighbor interactions. It also should be noted that the same quantity, the dipole matrix element, is responsible for the creation of the exciton by light, for the interaction with the neighbors, for the width of the band of states, and for the velocity of propagation.

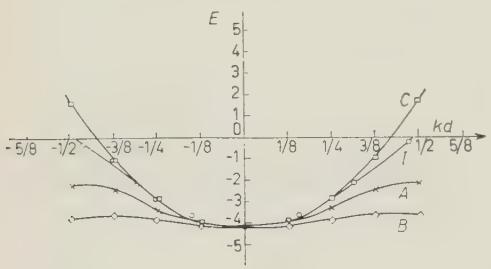


Fig. 9.—Dependence of energy on propagation vector $\boldsymbol{k}$ in a face-centered cubic crystal, for an exciton with dipole moment perpendicular to $\boldsymbol{k}$. The curve I results from an integral approximation and A, B, C are for $\boldsymbol{k}$ in the 100, the 110, and the 111 directions. (After Heller and Marcus, reference [15].)

On this basis, the energy spectrum of the excited states, $E_n(\boldsymbol{k})$, is found to vary quadratically with $\boldsymbol{k}$, near $\boldsymbol{k} = 0$. See Fig. 9. The coefficient, which is related to the effective mass for an exciton, depends on the orientation of the dipole moment of the exciton (primarily parallel to the electric vector if created by optical absorption) with respect to the propagation vector. For an optically excited exciton the effective mass is equal to

$$m^* \cong \frac{m_e}{f_n} \frac{R}{a_n}, \tag{II-5}$$

where R is the lattice constant, m_e the electronic mass, a_n is the spatial extent of the localized function ψ^n, and f_n the oscillator strength connecting the n-th state with the ground state. Thus in typical cases we may expect an effective mass for the exciton to be nearly as small as the electronic mass, if its creation in an optical transition is allowed. This implies a width to such an exciton band of the order of one eV. Near the bottom of the band a velocity may be assigned to the exciton by forming a wave packet in the usual way, and the velocity $(1/\hbar)\nabla_k E(\boldsymbol{k})$ is of the order $(\hbar/m_e)f_n(a_n/R)\boldsymbol{k}$.

2. – Selection rules.

The optical properties associated with excitons are expressed in terms of the matrix element of the Hamiltonian describing the electromagnetic field, Eq. (I-5) [16]. The selection rule for the conservation of energy appears in the usual way, stating that a transition is unlikely unless

$$E_n(\boldsymbol{k}) = E_0 + \hbar\omega \,, \tag{II-6}$$

where $\hbar\omega$ is the energy of the incident photon of momentum $\hbar\boldsymbol{\eta}$. Computing the matrix element of Eq. (I-5) with the wave functions (II-1) and (II-4) we obtain

$$\langle n, \boldsymbol{k} | H_R | 0 \rangle = N^{-\frac{1}{2}} \frac{ie\hbar A_0}{mc} \sum_L \int \exp[-i\boldsymbol{k}\cdot\boldsymbol{R}_L] \Psi_L^{n*} \sum_i \boldsymbol{a}\cdot\nabla_i \exp[i\boldsymbol{\eta}\cdot\boldsymbol{r}_i] \Psi^0 \,. \tag{II-7}$$

The integral is zero unless $i = L$, in which case we obtain

$$\langle n, \boldsymbol{k} | H_R | 0 \rangle = N^{-\frac{1}{2}} \frac{ie\hbar A_0}{mc} \sum_L \exp[i(\boldsymbol{\eta} - \boldsymbol{k})\cdot\boldsymbol{R}_L] \int \psi_L^{n*} \boldsymbol{a}\cdot\nabla_L \psi_L^0 \, d\tau \,, \tag{II-8}$$

and since the atoms are identical, the integral is independent of L. This integral, $\langle n | \boldsymbol{a}\cdot\nabla | 0 \rangle$, is simply the atomic matrix element of $\boldsymbol{a}\cdot\nabla$ which may be related to the dipole matrix element $\boldsymbol{r}_n$ through the well known theorem

$$\langle n | \boldsymbol{a}\cdot\nabla | 0 \rangle = \frac{-m}{\hbar^2} (\varepsilon_n - \varepsilon_0) \boldsymbol{a}\cdot\boldsymbol{r}_n \,. \tag{II-9}$$

Except for surface effects, the summation in (II-8) vanishes if $\boldsymbol{\eta} \neq \boldsymbol{k}$, and gives N if $\boldsymbol{\eta} = \boldsymbol{k}$. This is the momentum selection rule, which asserts that the exciton takes up the momentum of the photon in the absorption process.

The atomic symmetries are seen to enter into the computation of (II-9). If the atomic transition is allowed, the corresponding transitions in the crystal

are allowed. This is something of an oversimplification, but we will not carry it further at this time. The magnitude of the dipole matrix element in (II-9) determines the strength of the absorption band, and also the width of the band of excitation states, the effective mass, etc. If in the crystal the excited state wave function is very large compared with the ground state, the dipole matrix element will be reduced by the poor fit, and relatively little absorption will result.

Observable optical quantities are proportional to the absolute square of the matrix element in (II-8),

$$|\langle n, \boldsymbol{k}|H_R|0\rangle|^2 = \frac{e^2A_0^2}{\hbar^2c^2}(\varepsilon_n - \varepsilon_0)^2(\boldsymbol{a}\cdot\boldsymbol{r}_n)^2[N\delta_{\boldsymbol{k},\boldsymbol{\eta}}]\,. \tag{II-10}$$

Except for the factor in square brackets, this is the same value we would have obtained for an isolated atom. We see that for any given band of excitation states we have N times as much absorption as we would have from the corresponding level of a single atom; further, all of the transition probability occurs to a single state, for which $\boldsymbol{k} = \boldsymbol{\eta}$. It is this property which has led to the common remark that the optical properties associated with excitons are the same as for isolated atoms.

3. – Coherence.

It has not been generally noticed that one striking dissimilarity exists. Just as in absorption, where N atoms absorb N times as well as does one, so in emission the system emits with N times the spontaneous emission probability of one atom. This point is intimately connected with the question of coherence, and has received inadequate attention. It has been stated by a number of workers that emission times for excitons are of the order of or longer than typical atomic lifetimes. However, we see that in a periodic potential the lifetime will be extremely short from the one state which can be prepared by optical excitation, i.e., the state for which momentum and energy are conserved. Indeed, if absorption of a photon prepared the $\boldsymbol{k} = \boldsymbol{\eta}$ state, emission would occur in times short compared with a lattice vibration, so that the exciton would disappear before it could be scattered to another state. In the case of perfect incoherence the lifetime would be as long as an atomic lifetime; in this case we could say formally that the probability of occupation of the $\boldsymbol{k} = \boldsymbol{\eta}$ state is $1/N$, that the probability of emission from that state is N times an atomic probability, so that the resultant emission probability is the same as for an isolated atom. (Alternatively, we could state on simple physical grounds that for perfect incoherence the atoms are emitting

as independent atoms.) In the case of some preferential occupation of the $k = \eta$ state, as would be expected if the $k = 0$ state is at the bottom of the band, the emission probability would be greater than the atomic value, but not N times as great.

Several complications arise in attempting to assess the implications of this theorem, for example, the question of natural line width of the $k = \eta$ state if the emission probability from this state is very large. Also, there must exist some natural « saturation » mechanism to prevent emission probabilities from achieving unreasonably high values. These questions have not been answered, and surely deserve theoretical attention. Of one thing we may be certain, however, and that is that a simple statement of the momentum conservation law $k = \eta$ for absorption and emission does not tell the whole story by any means.

Optical absorption measurements have been made on many organic and inorganic crystals. Up to the present I think it safe to say that not a great deal has been learned from them, the theoretical work required for their detailed interpretation usually not having been carried out. In general, the calculated energies of the absorption bands are in at least vague agreement with experiment. The strengths of the transitions, where they are known, are in somewhat worse agreement. The widths and shapes of the absorption lines have almost invariably not been computed, nor is the temperature dependence understood. Fine structure in absorption has been considered in some detail in a few organic [17] and ionic [18] crystals, and is beginning to be treated in the alkali halides [19]. By-and-large, we know that absorption lines of varying width do occur in many insulating crystals, but in most cases that is about all we know. It would be useful to have more clear-cut information about the photo-conductivity (or its absence) associated with these transitions. Fig. 10-14 show some typical absorption curves for ionic and organic crystals. Most of the work with very fine absorption lines is published only as spectrographs, without microphotometer tracings, and cannot conveniently be shown here.

In a few specific crystals of aromatic molecules the transitions have been identified in some detail. Even in these

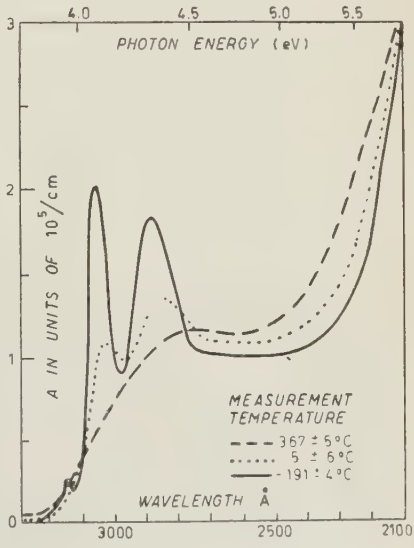


Fig. 10. – Optical absorption in annealed BaO films. (After Zollweg, reference [18].) The peaks tend to disappear with increasing temperature.

few cases an understanding of line breadth cannot be said to exist. Certainly we cannot yet answer the questions about coherence from the optical absorp-

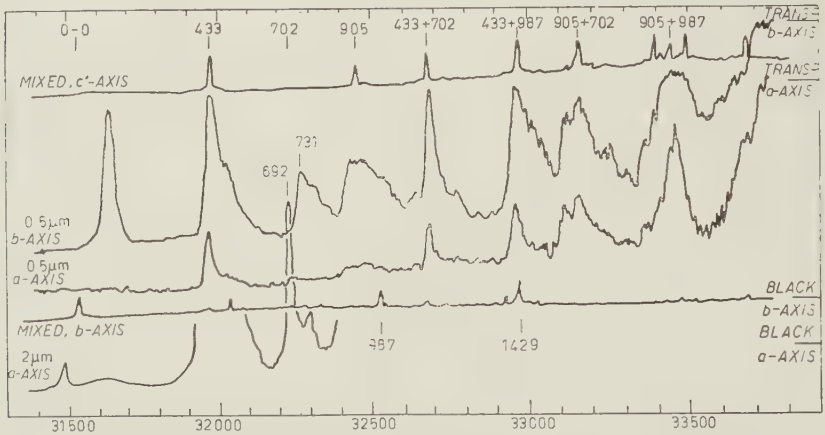


Fig. 11. – Microphotometer tracings of the absorption spectra of naphtalene at 20 °K. The different curves are for pure and mixed crystals with different orientations of polarization. (After McClure and Schnepp, reference [17].)

tion measurements. In crystals of the aromatic molecules which have more than one molecule per unit cell, differing orientations occuring for the various

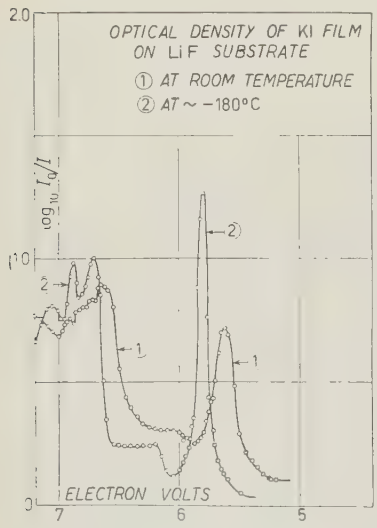


Fig. 12. – Optical absorption in evaporated KI film. (After Teegarden, reference [19].)

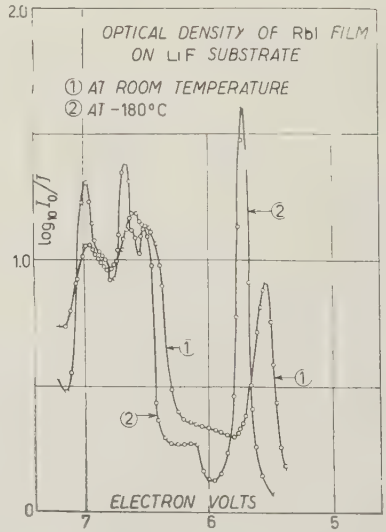


Fig. 13. – Optical absorption in evaporated RbI film. (After Teegarden, reference [19].) Note the similarity to the KI curves, and the additional structure.

molecules, more than one transition occurs for a given molecular state. This effect is generally known as « Davydov splitting » [20]. The theory of Davydov splitting is based on coherent excitation, and its occurence implies at least short-range coherence over a few molecules.

Emission measurements are less commonly made. Only in the past year or so have the alkali halides been known to luminesce with any appreciable efficiency without intentional doping [21]. In this case a large Stokes' shift occurs, demonstrating that if an exciton is created by optical absorption, something drastic happens to it before it emits. At temperatures above liquid nitrogen temperatures, the luminescence efficiency decreases rapidly for NaI and KI, and nitrogen apparently is not cold enough for KBr luminescence. (TEEGARDEN, private communication.) In organic crystals more information is available and Stokes' shifts are smaller. Decay times and quantum yields have not usually been measured. Considerably less is known about emission from excitation states, even, than about absorption to these states.

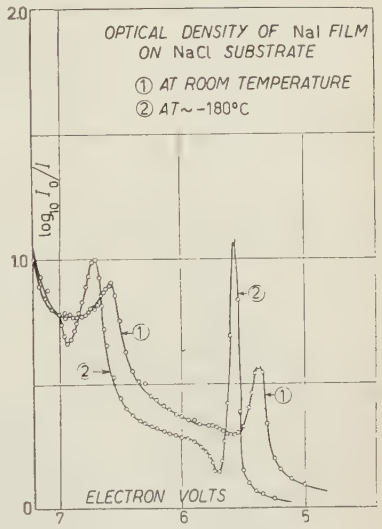


Fig. 14. – Optical absorption in evaporated NaI film. (After TEEGARDEN, reference [19].) The structure has partially disappeared with this lighter cation, as though spin orbit coupling on the cations might be responsible for the splitting of exciton levels.

Some optical experiments have been proposed which may assist in our understanding of the problem of coherence. FOX and YATSIV [22] have pointed out certain long-range interactions in organic crystals which should influence the energies and polarizations of the transitions as a function of the *direction* of propagation of the exciting light, if the region of coherence is large with respect to the wavelength. Experiments with varying polarization and direction of incidence should allow a demonstration of this coherence, if it exists.

Another possible experiment is suggested by the dependence of the emission probability on the size of the coherently excited region. The experiment could best be performed using a thin crystal of a material in which the lowest energy atomic transition has an oscillator strength of the order of or less than 10^{-4}. (Where the transition is strong, absorption itself limits the possible coherence in the direction of propagation.) With an oscillator strength of 10^{-4}, one would expect an atomic decay time of about 10^{-4} s. A deviation from this value could give information about the amount of coherence.

FOX [23] has also suggested experiments on organic polymers as a means

of building « small excitons ». In this case one has control of the maximum possible extent of coherent excitation.

In summary of the present experimental information, let us note that there exists no demonstration of long-range coherence of excitons in any crystal. In part this is because the importance of coherence has not been well-understood. As we have seen, coherence *is* important in affecting the optical properties; it is also intimately connected with the question of propagation, or of energy transfer. If the system is not coherent over at least the order of a wavelength, it becomes meaningless to describe the state with a propagation vector. Energy transfer may still take place, but a conceptual change is required.

4. – Propagation and energy transfer.

If a reasonable degree of coherence exists, and we ignore emission from the excited state, we may consider the exciton to travel through the crystal with a definite momentum, being scattered occasionally to a new momentum, and eventually to disappear, perhaps by transferring its energy to some imperfection. If it is created in the $\boldsymbol{k} = \boldsymbol{\eta}$ state, it starts off with a velocity $\hbar\boldsymbol{\eta}/m^*$ or about 10^5 cm/s with a large oscillator strength. This corresponds to a transit time between adjacent unit cells of about 10^{-13} s. During its scattering by lattice vibrations it becomes thermalized, and may reach a velocity 10^7 cm/s, corresponding to a transit time 10^{-15} s. It is not known if this ever happens in practice. If the system is not largely coherent we may not compute a velocity in this way.

Rather than assume coherence, and specify a momentum, we could assume incoherence and specify the position of the excited atom or molecule at a given instant, as in Eq. (II-2). Then we may compute the probability for the excitation energy to jump to another specified site. This is the approximation used to calculate energy transfer in the study of sensitized luminescence [24], where the energy absorbed on one species of atom is transferred to another species where emission occurs. It is the same mechanism which is responsible for transfer, that is, the dipole-dipole interaction, but the calculation of the transfer probability proceeds in a different way. Resonance transfer to a near neighbor may occur as rapidly as 10^{-13} s in an ideal case, or very much slower in other cases. Unless both the atoms involved have allowed optical transitions in the energy region of interest, the transfer occurs by other means, e.g., exchange or dipole-quadrupole interaction.

Transfer to an impurity atom called the activator (or emitter), has been demonstrated in a number of systems, in which the absorbing species (sensitizer) is either the host lattice or a second impurity. Reasonable agreement

between theory and experiment has been generally observed. In some cases of host-sensitized luminescence there has been definite evidence of energy transfer between one host-crystal ion and another [24*b*, 25]. In the observed cases the transition moment is weak, so that transfer tends to be slow, and without the activator present the systems luminescence with a large Stokes' shift, indicating still slower transfer. Indeed, there can be little doubt that the energy migrates with an effective mass of the order of nuclear masses. In this case essentially perfect incoherence doubtless prevails, representing one extreme of energy propagation.

The alkali halides probably represent other examples toward this extreme. Calculations by LEURGANS [26] show that the scattering by lattice vibrations reduces the mean free path to only a few atomic distances, and the large Stokes' shift [21] indicates that the energy is localized, i.e., trapped, when emission occurs. The host-sensitized luminescence experiments on several tungstates, molybdates and phosphates, referred to in the last paragraph, should be repeated with the alkali halides when suitable activators are found. These experiments are capable of furnishing a diffusion length for the exciton. An impurity halide ion, such as a small amount of NaI in NaCl, may prove to be a satisfactory detector for the exciton.

Examples toward the other extreme may be some of the organic crystals, where Stokes' shifts are small and transfer of energy to impurity activators is probable [17*b*]. Further experiments on the variation of transfer efficiency with impurity concentration would be valuable.

Other experiments demonstrating energy transfer have been performed in the alkali halides. One of the first, by SMAKULA, consisted in the irradiation of the long wavelength absorption tail [27]. See Fig. 15. It was found that this procedure resulted in the creation of *F*-centers, approximately one *F*-center per absorbed uv photon in the initial stages. Thus apparently each (« unusual ») exciton migrated to a negative ion vacancy and dissociated, leaving its electron bound to the negative ion vacancy, while the positive hole diffused away. The qualification « unusual » is made because of the circumstance that photon

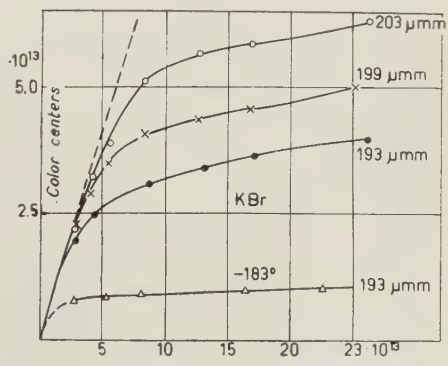


Fig. 15. – Coloration of KBr crystals upon irradiation in the long wavelength absorption tail. The initial slope indicates that almost every absorbed uv photon results in the creation of an *F*-center until a saturation effect occurs, such as the filling of all available negative ion vacancies or the ionization of the *F*-centers so produced (After SMAKULA, reference [27].)

energies well out on the tail of the absorption band were used, so that one is not certain that the resulting processes were characteristic of the « average » exciton produced at the peak of the first fundamental absorption bands. A similar remark may be made in reference to existing work on the effect of low-energy excitation on the ionizing of F-centers and thus producing an internal photocurrent [28].

A series of experiments by Apker, Taft and co-workers has shown that absorption of light in the first fundamental absorption bands of the alkali iodides results not only in the creation of F-centers, but also in their ionization [29]. The ionized electrons are detected as an external photocurrent. The efficiency of this process was found to be quite high, of the order of one electron ejected per one hundred incident uv photons near the peak of the first fundamenalt absorption band. See Fig. 16. These experiments demonstrate clearly that energy is transferred from the excitons produced by absorption, both in the creation of F-centers and in the ionization of them. The concentrations of F-centers were sufficiently high, however, that one cannot conclude that the exciton necessarily migrated any distance before finding an F-center; the experiments can also be interpreted on the basis of the sensitized luminescence process, in which case the exciton, even though localized, might transfer energy to an F-center up to a dozen lattice constants distant. The whole of the evidence, I believe, suggests that some migration occurs, though not necessarily very much in the ionic crystal, where interaction with lattice vibrations and static imperfections is so great.

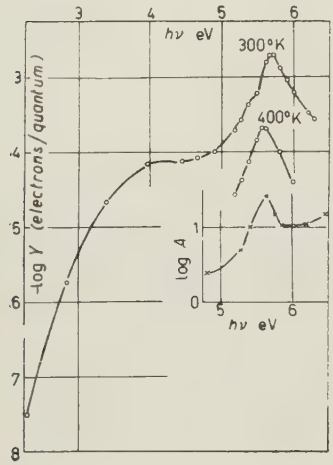


Fig. 16. – Logarithm of the external photoelectric yield *versus* photon energy in a KI film. The yield is defined as the ratio of the number of external photoelectrons to the number of *incident* photons. Since the film is thin the yield tends to follow the absorption, shown in the inset. (After Apker and Taft, reference [29].)

Other evidence for energy migration occurs in the sensitization, desensitization, and supersensitization of photographic emulsions [30]. It is found that suitable organic dye molecules, present as a thin film covering each silver halide grain in the emulsion, have a profound effect on the efficiency of production of the photographic latent image. The several effects have been best interpreted on the basis of absorption of the incident photons by the sensitizing dye molecules, and migration of the energy among themselves until

the energy is transferred to the silver halide grain. Supersensitizers are then regarded as particularly efficient sites for the transfer to the grain, desensitizers as traps for the energy [30].

One of the most striking indications of energy migration occurs in a recent experiment by BALKANSKI [31]. He has found in cadmium sulfide that a photocurrent can be produced at one end of a crystal upon irradiation with light at the other end. Considerable ingenuity was exercised to ensure that the effect did not arise from a trivial cause. For example, the measured time delay for the effect demonstrates that it is not a result of light-scattering. An electric field applied between the two ends of the crystal has an effect on one component of the induced photocurrent, as if electrons were migrating down the crystal, but another large component is field-independent. Many other measurements were made, strongly indicating that excitons produced by the light migrate a distance of millimeters along the crystal, eventually to dissociate at a surface into a free electron and hole, so that they can contribute to the current. All of his observations seem consistent with this interpretation, and although other explanations are conceivable, none that has been forthcoming seems less unlikely than the simple one here discussed. This seems to me to be a highly important experiment, being really clear-cut evidence for macroscopic migration of electronic excitation energy in a solid without the involvement of charged particles. Further experiments along this line will surely be awaited with great interest.

Another experiment on CdS by REYNOLDS *et al.* may lend support to this interpretation [32]. With a rectifying barrier on one end of the crystal, so that only the arrival of positive holes could contribute to a steady photocurrent, REYNOLDS induced a current by irradiation of the other end of the crystal. He also found a diffusion length of the order of millimeters, and his observations seem explicable on the basis of the diffusion of excitons down to the barrier where dissociation occurs. Although inherently not as many checks are possible on this experiment, this may prove to be another observation of macroscopic energy migration.

Professor BUSCH in his lectures mentioned another experiment on the thermal conductivity of semiconductors which may prove to be further evidence for the migration of excitons.

III.

Isolated Imperfections.

In this lecture we shall be concerned with the optical properties of a system specifically associated with the presence of a low concentration of imperfections randomly dispersed in the host crystal. Our intent is primarily to discuss

the various mechanisms which occur, and not to give a thorough summary of either the experimental or theoretical work [33].

Throughout it will be assumed that the optical transitions of interest occur in an energy region for which the host crystal is normally transparent. In this case the impurities or other imperfections may be treated as isolated, since they are not in resonance with the atoms of the host crystal, and since they are well-separated, on the average, from each other. This assumption, which is valid in many realizable cases, allows a considerable simplification, as we shall see.

For simplicity let us refer to these isolated imperfections as impurities, whether they are impurities, color centers, or other imperfections. The impurity plus its surroundings will be called a center. Our purpose is to understand the absorption and emission resulting from the presence of these centers. The kind of question one naturally asks is the energy of the absorption and emission bands, the absorption cross-section, decay time, and quantum yield for luminescence, the width and shape of these bands, and the temperature and pressure dependence of all of these quantities. Although a complete answer can by no means be given at the present time to these questions, our understanding is surely greater in this area than it is for exciton and band-to-band transitions.

A relatively simple model is available to us. We consider an isolated center as qualitatively similar to an atom or molecule *in vacuo*. Macroscopic corrections are introduced for the influence of the surrounding medium upon the electromagnetic field with which the center interacts, and microscopic account is taken of the interaction between the impurity and its immediate surroundings. That is, we make use of known bulk properties of the host crystal, such as index of refraction, compressibility, and thermal expansion to describe the macroscopic properties; microscopic interactions must be described in terms of detailed knowledge of the atoms of the host crystal and the imperfection itself.

Two opposite extremes may be made use of in the calculation of the interaction of the impurity with its surroundings. On the one hand, when the electronic wave function of the impurity is large with respect to the unit cell, the wave functions and energy levels are computed in the static lattice with the lattice constant characteristic of a given temperature; then the interaction with the vibrating surroundings is calculated as though the same vibrational spectrum existed in the neighborhood as would occur in the perfect crystal. This is evidently a good approximation when all properties of the impurity are similar to those of the atom or molecule which it replaces. It is not quite as evident that this procedure is valid merely because the electronic wave functions of the impurity are spatially extended, since the « hardness » of the core may influence the local lattice vibrations. Even in this case, however,

if the electronic wave function is extensive, it may « sample » the lattice vibrations over a large region, and tend to minimize the importance of local deviations from the normal lattice vibrational spectrum. Calculations have been made in this approximation by LAX and BURSTEIN for boron impurities in silicon, where reasonable success has evidently been achieved [34]. The approximation does not seem as satisfactory for F-centers in the alkali halides, where the wave function is more closely localized. Reference [35] contains a list of some of the theoretical papers on this topic which use this approximation.

We shall be primarily concerned here with the other extreme, in which the significant wave functions for the impurity are well-localized, and in which the interactions of the impurity with its neighbors are appreciably different from those of the host crystal atom which it replaces. In such a case, the interaction energy and hence the frequencies of vibration of the immediate neighbors of the impurity may differ considerably from those of the bulk crystal. In this event, transition energies and interactions with lattice vibrations may be considered on the basis of the local interactions alone, and the bulk spectrum of lattice vibrations may be ignored. Another statement of this conclusion has been formulated by KRUMHANSL and BJORK [36]. They have discussed an « exclusion » theorem which asserts that if a *new*, *localized* vibrational frequency is introduced by the insertion of an impurity in the lattice, other modes of oscillation tend to be repelled from the region of localization. At high temperatures, where equipartition of energy obtains, this statement becomes self evident, since, because the localized mode has $\frac{1}{2}kT$ of energy, the presence of other modes of oscillation in the vicinity would increase the local temperature.

For localized centers of this type, VON HIPPEL [37] and SEITZ [38] introduced an important method of description about 20 years ago. They considered the description of these centers on a « configuration co-ordinate » (C.C.) scheme, on the basis of an analogy with molecular physics. Just as the potential energy of a hydrogen molecule may be plotted as a function of the inter-nuclear separation, they envisaged the potential energy of the center to be plotted as a function of the co-ordinates of the surrunding atoms. Different surfaces will of course be obtained for each electronic state of the center. In general this would lead to many-dimensional surfaces, in which the energy would be a function of as many co-ordinates as there are degrees of freedom among the atoms interacting with the impurity. The scheme is particularly useful when the number of co-ordinates which influence the *separation* of two surfaces is small. Degrees of freedom, i.e., configuration co-ordinates, for which the variation of the energy at the center is independent of electronic state need not then be treated in detail for most purposes, and we may concentrate our attention on the relatively few co-ordinates remaining. In the

case of a tightly bound impurity atom, we would expect the important interactions to be with nearest neighbors. Thus if there are 6 nearest neighbors, we might expect to need 18 C.C. to describe the center. However, it is likely that only radial motion of the neighbors would have a profound effect, transverse motion not changing the distance to the impurity to a first approximation. Thus we are reduced to 6 C.C. Of these 6 modes we would expect the « breathing » mode, in which all six atoms move in phase radially, to be the most important; the others, in which some atoms are moving in and others out at a given instant, would have less effect because of a tendency for cancellation. Thus we might expect to reduce the problem to one important co-ordinate and several others of less significance. Much work kas been done assuming just one significant co-ordinate. In addition Lax [35] has shown that a description in terms of one co-ordinate is frequently possible, even when many degrees of freedom are really involved. This co-ordinate would in general be a fictitious one, with possibly a temperature-dependent mass.

Near the positions of equilibrium of the surrounding atoms, small displacements have a quadratic effect on the potential energy of the center. Thus for small displacements the center behaves as a harmonic oscillator, with the usual vibrational energy levels and wave functions. For larger displacements one would expect deviations from parabolic dependence on each C.C. When the neighbors approach the impurity the energy should rise faster than quadratically because of the exponentially rising repulsive energy.

The equilibrium positions of the surrounding atoms will in general depend on the electronic state of the impurity, since changing its charge distribution changes the interactions with the neighbors. Thus the minima of the various energy surfaces occur at different values of the configuration co-ordinates. In a qualitative fashion one may predict the direction of the shift in ionic crystals: if the impurity is a positive ion, raising it to an excited state extends its electronic charge distribution further away from its nucleus. Thus the positive ion is effectively made more positive and has a greater attraction for the nearest (negative) ions. Hence we would expect the minimum in the excited state to occur at smaller inter-ionic distances than the equilibrium values in the ground state. When the impurity is a negative ion or an *F*-center, the arguments are reversed. These conclusions seem to be in agreement with the small amount of existing experimental information.

Similarly, near the minima, the curvatures will differ for various electronic states. It seems likely that in general the excited state curves would have less curvature near their minima than that for the ground state, because of their more diffuse charge distributions. This conclusion also seems to be generally borne out by experiment.

In the case of just one important C.C., we might obtain energy curves for two electronic states something like those in Fig. 17. For more than one C.C.,

this curve may be considered a cross-section of the energy surface, holding all but one C.C. fixed. A few of the vibrational levels are shown.

Let us now discuss the theory of optical properties of such centers. The wave functions are of the form $\Phi_{al}(\boldsymbol{r}, x)$, where the subscript a labels the electronic state and l the vibrational state, $\boldsymbol{r}$ stands for the co-ordinates of the electrons and x for the nuclear co-ordinates, that is, the C.C.

Several approximations are commonly made, with varying degrees of validity. One approximation, usually a good one, and almost always invoked in practice, is the Born-Oppenheimer, or adiabatic, approximation, which allows a separation of electronic and nuclear motion. Because of the disparity between electronic and nuclear masses, and the resulting disparity in electronic and nuclear velocities and frequencies, the electronic wave function at any given istant may be computed as if the (vibrating) nuclei were fixed in their positions at that instant. This approximation allows a separation of variables in the quantum mechanical formulation which greatly simplifies the problem. According to the Born-Oppenheimer approximation, we separate $\Phi_{al}(\boldsymbol{r}, x)$ into a product of two functions

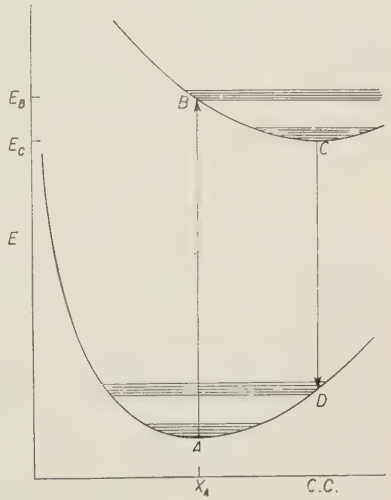


Fig. 17. – Schematic potential energy of a center as a function of a configuration co-ordinate. The equilibrium position of the center in the ground electronic state would be near A, in the excited state near C.

$$\Phi_{al}(\boldsymbol{r}, x) = \psi_{ax}(\boldsymbol{r})\, X_{al}(x)\,, \tag{III.1}$$

where $\psi_{ax}(\boldsymbol{r})$ is a function of the electronic co-ordinates alone, depending parametrically on the nuclear co-ordinates and of course on the electronic state, and $X_{al}(x)$ is a function of the nuclear co-ordinates alone, depending on the vibrational and the electronic state. Optical, symmetry-allowed transitions will be determinated by the dipole matrix element of the electronic co-ordinates $e\boldsymbol{r}$ between two electronic states, say an initial state a with vibrational states l and a final state b, with vibrational states m. The absorption coefficient is found to be

$$\sigma(E) = \left[\frac{n}{\varkappa_0}\left(\frac{\mathcal{E}_{\mathrm{eff}}}{\mathcal{E}}\right)^2\right]\frac{4\pi^2 E}{3\hbar c}\, S_{0c}(E)\,, \tag{III-2}$$

and the emission probability per unit time is given by

$$W(E) = \left[\frac{n^3}{\varkappa_0}\left(\frac{\mathcal{E}_{\mathrm{eff}}}{\mathcal{E}}\right)^2\right]\frac{4}{\hbar^4 c^3}\int E^3 S_{cl}(E)\,\mathrm{d}E\,. \tag{III-3}$$

The quantities in square brackets arise from the influence of the medium on the electromagnetic field with which the center interacts; n is the index of refraction, $\varkappa_0 \cong n^2$ the high frequency dielectric constant, and $(\mathcal{E}_{\text{eff}}/\mathcal{E})$ an effective field ratio. See LAX (ref. [35]), whose treatment we are generally following here, for the macroscopic origin of these factors, and ref. [39] for a partial treatment of the microscopic origin. The functions S_{ba} and $S_{ab}(E)$ are shape functions for the absorption and emission bands; each is of the form

$$\text{(III-4)} \qquad S_{ba}(E) = A v_l \sum_m \left| \int X^*_{bm}(x) \boldsymbol{M}_{bl}(x) X_{al}(x)\, \mathrm{d}x \right|^2 \delta(E_{bm} - E_{al} - E)\,,$$

where

$$\boldsymbol{M}_{bl}(x) = \int \psi^*_{bx}(\boldsymbol{r})\, e\boldsymbol{r}\, \psi_{al}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}\,,$$

is the dipole matrix element for the electronic transition. The symbol Av_l indicates an average over the initial vibrational states, weighted with the usual Boltzmann factor. $\sum_m$ indicates a sum over all final vibrational states. The δ-function is the Dirac δ-function, and expresses the conservation of energy in the transition.

A common approximation at this point (*) is the removal of $M_{ba}(x)$ from the integrand of (III-4). This is sometimes referred to as the Condon approximation. It is a valid one when the electronic dipole matrix element does not depend markedly on the oscillations of the nuclei. Very little is known about the validity of this procedure. One rough calculation has indicated that it is not a very good one in ionic crystals [40]. Nevertheless it is almost essential to follow this procedure in order to continue in a simple manner. We thus replace $\boldsymbol{M}$ by a suitable average and remove it from the integral, obtaining

$$\text{(III-5)} \qquad S_{ba}(E) = \overline{\boldsymbol{M}^2} A v_l \sum_m \left| \int X^*_{bm}(x)\, X_{al}(x) \mathrm{d}x \right|^2 \delta(E_{bm} - E_{al} - E)\,.$$

According to Eqs. (III-2) to (III-5), the absorption or emission should consist of lines, each corresponding to a transition involving two discrete vibrational levels. In practice, in ionic crystals, these lines are smeared out by other interactions, and a broad band of absorption or emission results. If we are not interested in accounting for possible fine structure, which is generally not observed, another simplification is possible which corresponds to the Franck-Condon principle.

Let us assume that the minima in Fig. 17 are well-removed, so that vertical

(*) It would be possible to put off this approximation until after one more step in the development, but since we shall have to make it anyway, we do it now for simplicity.

lines drawn from near the minimum of one curve intersect the other at considerable distances above *its* minimum. To anticipate our conclusions, we may say that involved in the optical transition are final vibrational states of large quantum number. According to the Correspondence Principle, then, we could treat these states classically, so that each vibrational wave function (of high quantum number) is large only near the classical turning point. In quantum mechanical language the integral in (III-5) gets sizeable contributions only from a region of x near the potential energy curve, because of the cancellation elsewhere resulting from the rapid oscillation of $X_{bm}(x)$. Thus, for a given value of x, the only vibrational state m which contributes to the integral is that for which the potential energy curve $E_b(x)$ is equal to E_{bm}. Thus we may replace E_{bm} in the argument of the δ-function by $E_b(x)$, rewrite the absolute square of the integral, and obtain

$$\text{(III-6)} \qquad S_{la}(E) = \overline{\boldsymbol{M}^2} A v_l \sum_m \int \mathrm{d}x \int \mathrm{d}x' \, X^*_{bm}(x) X_{al}(x) X_{bm}(x') X^*_{al}(x') \cdot$$

$$\cdot \delta(E_b(x) - E_{al} - E) =$$

$$= \overline{\boldsymbol{M}^2} A v_l \int \mathrm{d}x \int \mathrm{d}x' \Big(\sum_m X^*_{bm}(x) X_{bm}(x') X_{a'}(x) X^*_{al}(x') \Big) \cdot$$

$$\cdot \delta(E_b(x) - E_{al} - E)) =$$

$$= \overline{\boldsymbol{M}^2} A v_l \int |X_{al}(x)|^2 \, \delta(E_b(x) - E_{al} - E) \, \mathrm{d}x \,,$$

where we have made use of the closure property for a complete set of eigenfunctions

$$\text{(III-7)} \qquad \sum_m X^*_{bm}(x) \, X_{bm}(x') = \delta(x - x') \,.$$

We may evaluate the integral in (III-6) by means of the δ-function, first changing the variable from x to $E_b(x)$, and obtain

$$\text{(III-8)} \qquad S_{la}(E) = \overline{\boldsymbol{M}^2} A v_l |X_{al}(y)|^2 \Big/ \frac{\mathrm{d}E_b(x)}{\mathrm{d}x}\Big|_{x=y} \,,$$

where y is the value of x which satisfies the conservation of energy, $E_b(y) - E_{al} = E$. Thus the shape function for absorption (or emission) is given by a thermal average, over the initially occupied vibrational states, of the initial probability of occupation of a given value of the C.C. times a Jacobian expressing the energy of the transition at that value of the C.C.

Lax has suggested a futher simplification to Eq. (III-6). He suggests that if a great many initial vibrational states l are involved in the transition, a mean value approximation may be used for E_{al}, i.e., $E_{al} \to E_a(x)$, and

we obtain,

$$S_{ba}(E) = \overline{M^2}\left[\frac{\mathrm{d}}{\mathrm{d}x}\left(E_b(x) - E_a(x)\right)_{x=z}\right]^{-1} A\,\nu_l\,|X_a(z)|^2, \tag{III-9}$$

where z is the value of the C.C. which satisfies conservation of energy $E_b(z) - E_a(z) = E$. The final average is the statistical mechanical probability for the occupation of the various values of the C.C. If this probability is computed with quantum (classical) statistics, the resulting expressions for σ and w are called the semi-classical (classical) Franck-Condon Principle. Either (III-8) or (III-9) gives rise to a familiar statement to the effect that optical transitions are unlikely when the positions of the nuclei change. That is, we treat only vertical transitions on the C.C. diagram. This is really a consequence of the cancellation of the matrix element, the integral in Eq. (III-5), for values of x away from the classical turning point.

Let us refer again to Fig. 17, in which we may trace some familiar processes. Suppose that the crystal is at low temperature, so that the center is in its zero point vibrational state, or very near to it, in the vicinity of A on the diagram. Absorption occurs, causing transitions to the neighborhood of B, where the vibrational state is high, inconsistent with the crystal's temperature. Violent oscillations ensue, in which heat energy is rapidly transmitted to the surroundings. Eventually the center cools off, reaching the vicinity of C in times of the order 10^{-10} or 10^{-11} s. Emission occurs some time later, say in 10^{-8} s, to a region near D, and again heat energy is transferred to the crystal, the center finally approaching point A. The reason for the Stokes' shift is apparent from this diagram; the difference between the energy of absorption AB and that of emission CD is converted into heat. This is quite a general phenomenon when the two minima are not directly above each other.

The strength of the transition, proportional to the integrated absorption cross-section or to the emission probability, is given correctly by all of the approximations made above. Indeed, assuming the Condon approximation, it is a constant, independent of variations in the width of the absorption or emission band through temperature or pressure effects. Thus the variation of the integrated absorption coefficient with changes in temperature would be sufficient (but not necessary) to demonstrate the failure of the Condon approximation.

A change in the maximum of the absorption or emission curve results from thermal expansion, which may be related to the effects of pressure [41], and from more specific temperature effects. Let us ignore for the present the effects of thermal expansion, which usually result in a shift of the maxima to lower energy. In the semi-classical approximation (III-9), the only temperature dependence occurs in the distribution function, which broadens with increasing temperature, but does not change its midpoint. Thus in this approximation

the maxima of the absorption and emission curves do not change. The use of the more accurate Eq. (III-8) does lead to a small shift, and we will return to this topic later.

Let us turn to a discussion of band shapes, which in (III-9) are given in terms of the distribution function and the Jacobian. A common approximation is to assume that over the range of C.C. involved in the absorption act, that is, in the vicinity of x_A in Fig. 17, the curve $E_b(x)$ is essentially a straight line of the same slope it has at $x_B = x_A$. Since $E_a(x)$ is essentially a constant near x_A, the Jacobian reduces to a constant, and the shape is determined solely by the distribution function. Now the thermally averaged, quantum mechanical distribution function is a Gaussian in x. Since there is a linear relation between x and the energy of the transition E, according to our assumption about the Jacobian, the absorption and emission bands should be Gaussians in energy at all temperatures. The width W of the Gaussian varies as the one-half power of the absolute temperature at high temperatures, and approaches a constant, determined by zero-point oscillations, at low temperatures. The functional dependence is given by

$$W \propto [\coth (\hbar\omega_a/2kT)]^{\frac{1}{2}}, \tag{III-10}$$

where $\hbar\omega_a$ is the separation of adjacent vibrational levels in the initial electronic state a.

As we have seen, the emission probability w, in a first approximation, should be temperature independent. It is observed, though, that the luminescence yield in most systems decreases with increasing temperature. This phenomenon, known as thermal quenching, has been interpreted with C.C. curves as follows [42]. Suppose that a center has been excited by the absorption of light, has emitted phonons, and the center is near the point C on the C.C. curve in Fig. 18. If an unusually large amount of thermal energy should become available to it, the center may find itself momentarily in a high vibrational state X_{bQ} with a vibrational energy equal to $E_Q = E_E - E_G$. At a point such as E on the diagram, a changeover to the ground electronic state can occur with conservation of energy, the electronic wave function changing to that of state a, and violent oscillations taking place characteristic of the vibra-

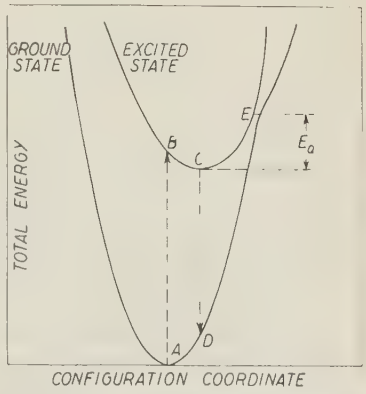


Fig. 18. Schematic potential energy diagram to indicate thermal quenching of luminescence according to Gurney and Mott [42]: (After Klick and Schulman, ref. [33].)

tional level X_{ak}, where $k \gg Q$. In the vicinity of E the vibrational wave functions X_{bQ} and X_{ak} are both large, so the transition can take place near E better than at other places, where at least one of the vibrational states is rapidly oscillating. After the transition the center would rapidly radiate phonons, eventually cooling off to the region A on the diagram. The probability of reaching the vibrational state Q is given by a Boltzmann factor $\exp[-E_Q/kT]$, and let us denote the probability of the transition, when the center has this energy, by the symbol s. Thus the probability of the quenching process is $P_Q = s \exp[-E_Q/kT]$, and the quantum yield for emission is the ratio

$$\eta = \frac{w}{w + P_Q} = \{1 + (s/w) \exp[-E_Q/kT]\}^{-1}, \tag{III-11}$$

where w is given by Eq. (III-3). This relation is at least qualitatively in agreement with experiment. The decay time would show a similar decrease with increasing temperature,

$$\tau = (w + P_Q)^{-1} = \eta/w . \tag{III-12}$$

Some centers do not seem to luminesce at all with any reasonable efficiency. Seitz [38] proposed a possible mechanism to explain such a phenomenon. If there is a great deal of relaxation of the center when it changes its electronic state, the surfaces for the two electronic states a and b may come very close or intersect without the necessity of thermal excitation required in the thermal quenching explanation. Fig. 19 shows such a case, where the center may jump back to the electronic state a when the center cools off to the minimum of the electronic state b. With a system described by this diagram, the center, when back in state a, may cool off to its absolute minimum, or to a quasi-stable position A. In either event little or no luminescence would be observed.

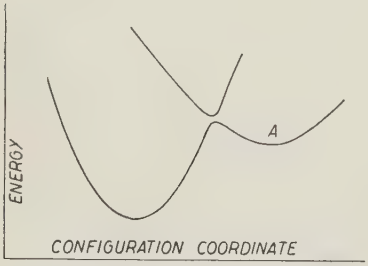


Fig. 19. – Schematic configuration co-ordinate curve. In such a center strong luminescence could not occur, except in the infrared. (After Seitz, reference [38].)

Another explanation of a possible absence of luminescence has been suggested [43], one which requires less lattice relaxation than the former. In Fig. 20, absorption raises the system to the vicinity of B, after which violent oscillations occur. If the surfaces for electronic states a and b intersect or come close to each other at some energy $E_C < E_B$, during the cooling process the center will reach a point such as C, where the non-radiative return to electronic state

a may take place, just as in thermal quenching. If such a transition is a likely one, a criterion for strong luminescence may be that the surfaces must not intersect or closely approach each other at energies less than E_B. Such a criterion leads, with a little algebra, to a prediction that the emission wavelength should not be more than twice the absorption wavelength, that is, that Stokes' shifts greater than a factor of two should not occur in simple centers. This condition seems to be generally but not absolutely satisfied.

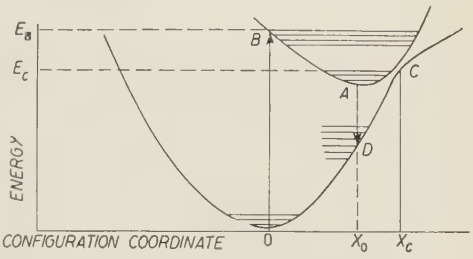


Fig. 20. – Schematic potential energy curve to explain possible absence of luminescence. In such a center a non-radiative transition can occur to the ground electronic state near point *C* on the diagram.

This same process may lead to auto-ionization, even at low temperatures, if the ionization levels dip below the point *B* reached in the absorption process. Such a case is shown in Fig. 21. Direct ionization would occur at the point *F*, but auto-ionization could occur while cooling from point *B*.

Even making use of the Born-Oppenheimer approximation, the Condon approximation, and the Correspondence Principle for the upper vibrational states, that is, using Eq. (III-8), an absolute calculation of the energies, strengths, and shapes of optical transitions is a difficult task. No serious attempt has yet been made on an absolute calculation, although the necessary procedure using symmetrically-orthogonalized wave functions to account for overlap [44], can be outlined in detail [45]. An alternative approach has been worked out by WILLIAMS [46], based on Seitz's interpretation of the thallous ion impurity in KCl [47]. From a group-theoretical analysis of the atomic wave functions and the crystalline symmetry, SEITZ proposed an energy level scheme and series of transitions for the KCl:Tl center, without, however, determining the energies and strengths quantitatively. WILLIAMS, by computing the interactions of the Tl^+ ion with its neighbors,

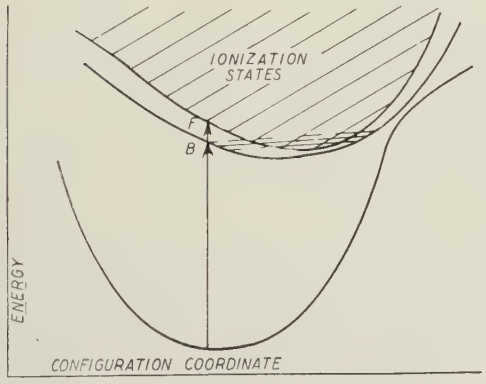


Fig. 21. – Configuration co-ordinate curve including ionization levels. Direct ionization can occur at energies above *F*; if the ionization levels dip below *B*, auto-ionization can occur while the excited center cools off after excitation to the discrete state.

which were assumed to oscillate in the « breathing » mode described earlier, calculated potential energy curves for the ground state $6s^2$ and an excited state $6s\ 6p$. He used as much experimental information as possible, to avoid inaccuracies associated with the use of the poor wave functions now available for Tl^+. The agreement with experiment, although not outstanding, is at least qualitatively good, some details naturally being more sensitive than others [48]. Some of the agreement may be fortuituous, and in some respects it is *ad hoc*; e.g., the variation with C.C. is arbitrarily chosen to be characteristic of Seitz's assignment of the $^3P_1$ state, whereas nothing in the calculation for the $6s\ 6p$ distinguishes between $^3P_1$ and $^1P_1$ levels. Application to Seitz's $^1P_1$ level results in poor agreement of the model with experiment. Nevertheless, although a number of criticisms can be made of the detailed procedure [45], this work is a pioneer (and still the best) attempt at the at least partially theoretical determination of C.C. curves.

Another technique, first introduced by Klick [49, 33], has been used successfully on a number of systems. Several experiments can be performed on a given system which may be used to determine the parameters needed to describe the C.C. curves for the ground and excited states. Such experiments are the low temperature absorption and emission peaks and band widths, and the temperaturature variation of the band widths. Indeed, the parameters can be overdetermined if enough experiments are performed, and the observed internal consistency is an indication that a one-parameter description on a C.C. model is a good approximation. It is important to note that the frequencies of vibration determined from experiment in ionic crystals are generally *not* the same as those of the pure host crystal, and further that they differ in ground and excited states. Thus the fundamental assumption in using this localized model seems to be justified.

Two additional effects, which have received little attention should also be mentioned here. One is the Jahn-Teller theorem [50], which was taken into account in Seitz's original KCl:Tl paper [47], but which has been largely ignored up to a year ago. The effect has recently been discussed in detail, with some emphasis on the solid state aspects, by Örik and Pryce [51]. The theorem states for our purposes that a geometrical arrangement of the atoms around a center in which the electronic state is degenerate is an unstable one, so that the atoms will move and destroy at least part of the degeneracy. (For our purposes the only exception occurs when the state is degenerate but the orbital state is not. Thus, the energy of the ground state of the F-center is the same regardless of the direction of the spin of the electron, but the orbital $1s$-like state is non-degenerate; this spin degeneracy would not be removed.) It is well-known that the degeneracy of p-states is not removed by a cubic field; according to this theorem, however, the surroundings will become distorted, there will no longer exist a cubic field, and the degeneracy, or part of it, will

be removed. Consider an F-center or a Tl^+ ion in an alkali halide; in the ground state the symmetry of its surroundings is cubic, and immediately after absorption of light the surroundings are still cubic. Afterward, lattice relaxation will occur which will remove at least part of the p-state degeneracy. The possibility arises, then, of observing polarization or structure in emission which would not be present in absorption. Little quantitative work has been done on the magnitude of the effect, but it is not large.

The other effect to be mentioned is the intersection and crossing of the C.C. surfaces describing different electronic states. In one-dimensional C.C. diagrams the curves are generally shown as not crossing, but, instead, as coming close and pushing each other away. The theoretical basis for deciding how the curves behave seems to rest on a theorem by von Neumann and Wigner [52]. They have shown on quite general grounds that, for states belonging to the same irreducible representation, crossing of the surfaces cannot occur unless at least three real parameters, e.g. C.C.'s, are allowed to vary. If the states belong to *different* irreducible representations, crossing can occur at will. Thus, if the ground state were an S state and the excited state a P-state, crossing could occur even on a one-dimensional plot. In the case of two excited P-states, as in KCl:Tl, for example, the surfaces could not cross in a one or two parameter description, and very likely not at all. A complete treatment of this topic does not seem to have been given.

Finally, I would like to discuss some calculated results on a specific, fictitious system, by Mr. N. N. Axelrod and myself. We were interested in various observations [53, 48] of the non-Gaussian nature of absorption and emission curves. We were led to investigate the effect of curvature of the final electronic state $E_b(x)$ in the Jacobian (III-8) and also the approximation leading from (III-8) to (III-9), namely the smearing-out of the initial vibrational levels. Accordingly we selected a specific system, in which the maximum absorption occurs to about the 40-th vibrational level of state b. At low temperatures where thermal occupation of all but the zero-point level is unlikely, we computed Eq. (III-8) with the exact Jacobian and also with the « straight-line approximation », in which the curve $E_b(x)$ is assumed to be a straight line of the slope of $E_b(x)$ directly above the minimum of $E_a(x)$. Let $x = 0$ be the minimum of $E_a(x)$, and let $x_0 > 0$ be the minimum of $E_b(x)$. With the straight line approximation we of course obtain a Gaussian distribution in energy. Two effects occur when the curvature is taken into account. First we note that high energy absorption occurs for negative values of x (since we assume $x_0 > 0$). Thus the increased slope of $E_b(x)$ where $x < 0$, as compared with the straight line approximation, results in more high energy absorption than the Gaussian. Similarly, there must be less low energy absorption. Hence there should appear a high energy tail in absorption as compared with a Gaussian. The same arguments apply in reverse for emission. Indeed, if the curvatures

of the two states a and b are the same, there should be precise mirror symmetry in the absorption and emission curves at all temperatures. If the curvatures are not the same, there should still be mirror-like symmetry at all temperatures, although the high energy tail in absorption might be more or less pronounced than the low energy tail in emission.

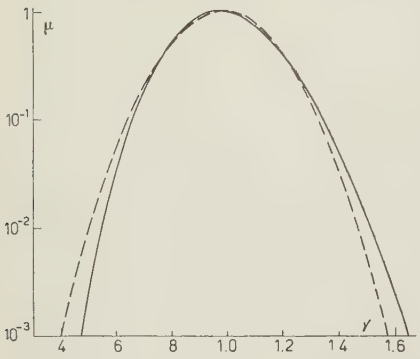


Fig. 22. – Calculated absorption from a fictitious center in which the peak absorption occurs to about the 40-th excited vibrational state. The curves are computed at low temperatures where kT is equal to one-tenth the separation of adjacent vibrational levels. The solid curve takes account of the curvature of the potential energy curve for the excited electronic state; the dashed curve is computed on the basis of the common assumption that the slope is constant, with the value it has directly above the minimum of the ground state. The abscissa is a reduced energy.

The other effect is that, because of the lesser slope for $x > 0$, a larger range of x contributes to a given range of energy in absorption. Thus the peak position of the absorption band is shifted to lower energy. To summarize, the peak is moved to lower energy, there is more high energy absorption, and less low-energy absorption than would be the case with the straight line approximation. Fig. 22 shows the computed curves for low temperature, where $kT = (1/10)\hbar\omega_a$, and these qualitative conclusions are demonstrated to be correct. The abscissa is a reduced energy of absorption, the solid curve is that obtained including curvature, and the dashed curve that obtained with the straight line approximation.

With a Gaussian curve the absorption coefficient is given by $\mu = \mu_0 \exp[-\alpha^2(E - E_0)^2]$, where E_0 is the center of the absorption band. Thus $[\log \mu_0/\mu]^{\frac{1}{2}} = \alpha(E - E_0)$ and a plot of this quantity should give a symmetrical V-shaped curve. Fig. 23 shows experimental data by Russell and Klick [53] on absorption by the F-center in KBr at 4 °K. Careful inspection shows that these data can be represented by a V-shaped curve where the V is slightly tilted in a clockwise direction, with a low energy slope about 1.2 times the high energy slope. In emission, according to the mirror symmetry arguments, the low-energy slope sould be somewhat *less* than the high-energy slope. For KCl:Tl at 4 °K Patterson and Klick [48] found a ratio about 1/1.1. In Fig. 24 is a plot of $(\log \mu_0/\mu)^{\frac{1}{2}}$ for our calculated case corresponding to Fig. 22. It is seen that the tilt of the V is qualitatively similar to Fig. 23, and it seems likely that the « tilted-V effect » is largely a result of the curvature of the final electronic state's C.C. curve. Thus, the smaller the separation in C.C.

of the minima of the two electronic states, the larger the effect must be. That is, if the center of the absorption or emission band involves the 40-th vibrational state, as assumed in our hypothetical center here, the curvature effect must be appreciably less than if we had treated a center where the 20-th vibrational level was responsible for the central transition. It is also necessary to recall that the curvature must generally increase, for either electronic state, as the co-ordinates move away from their equilibrium positions, because of the exponentially increasing repulsive energy. This would enhance the tilted-V effect.

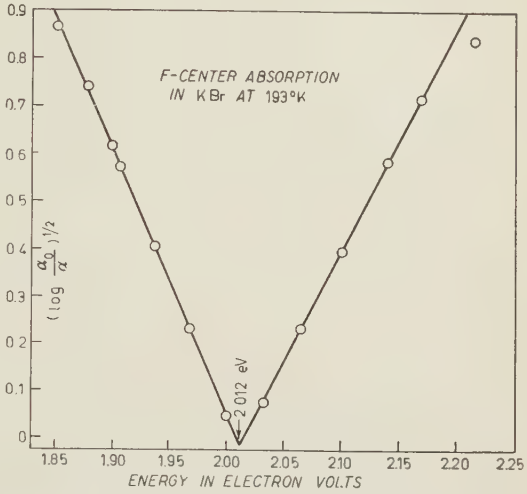


Fig. 23. – Experimental absorption curve for the F-center in KBr at 4 °K. Note the tilt in the V-shaped curve. (After Russell and Klick, reference [53].)

The other effect, that of the thermal occupation of higher vibrational levels, seems to have an exactly opposite effect. In absorption, the occupation of the higher vibrational levels raises the energy of the center toward the excited state curve, so that absorption at low-energy is enhanced and at high-energy is decreased, as compared with a Gaussian. Secondly, the maximum tends to be shifted toward *higher* energy, a somewhat surprising effect at first glance. Inspection of the schematic diagram in Fig. 25 makes this result plausible. The curve labelled 0 is the contribution to absorption from the ground vibrational state, and 1 from the next most important, for which $l = 1$. The center of the $l = 1$ curve is shifted by an amount $\hbar\omega_a$ toward lower energies from the center of the $l = 0$ curve. Thus the sum of these curves has its maximum shifted to higher energies with increasing temperature. It is not obvious that

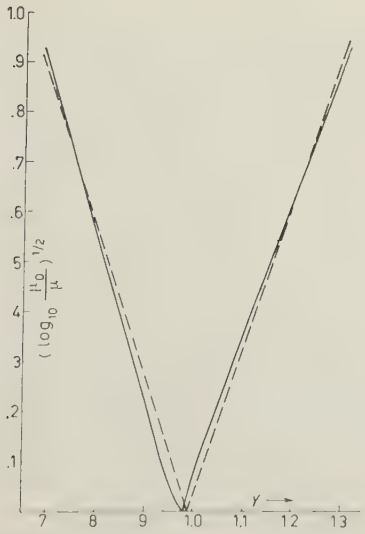


Fig. 24. – A different plot of the curve in Fig. 22. The dashed line results from the usual approximation of constant slope and the solid line from the more nearly exact treatment.

this will always be the case, though it seems likely at low temperatures. At a temperature $kT = \hbar\omega_a$ the calculated absorption curve using the more accurate Eq. (III-8) is shown in Fig. 26 in a solid line, with the result from

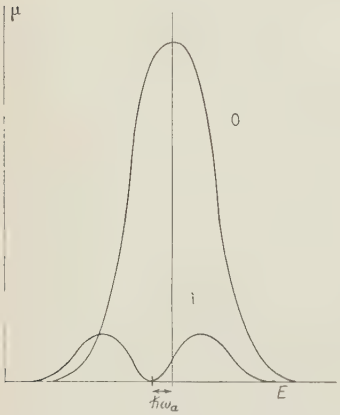


Fig. 25. – Contributions to the absorption coefficient from the zero-point and first vibrational level (schematic). Inclusion of the effects of the first excited level shifts the maximum of the sum to higher energy.

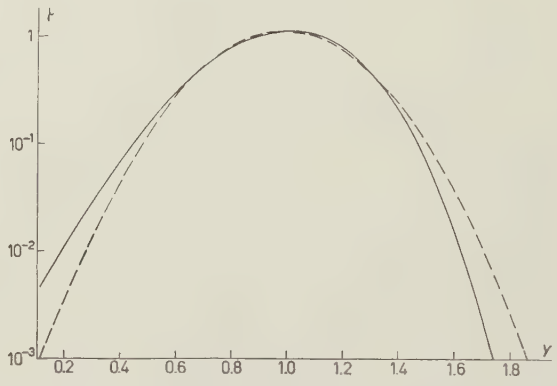


Fig. 26. – Calculated absorption coefficient for the center of Fig. 22 and 24 at a temperature where kT is equal to the separation of vibrational levels. The solid curve results from an exact treatment of the effects of the thermal population of higher vibrational levels; the dashed curve results from the use of an « effective temperature » in determining the thermally averaged distribution in energy.

Eq. (III-9) shown for comparison. The straight line approximation is made in both cases. Our qualitative arguments above are borne out here. We have also treated the curvature and temperature effect simultaneously for several temperatures. At the temperatures $kT = \hbar\omega_a$ the two effects almost cancel; i.e., the effect of treating curvature exactly is almost cancelled by the effect of treating exactly the thermal occupation of higher levels. Thus the Gaussian and the « exact » curve are almost indistinguishable. This will not always be the case; such a result depends on the details of the C.C. curves. However, it does appear that there will in general be a tendency toward cancellation, so that the commonly used semi-classical treatment with straight line approximation is better than it has a right to be.

* * *

It is a pleasure to acknowledge the assistance of many colleagues at the University of Rochester, without whose co-operation these lectures could not have been prepared in time.

REFERENCES

[1] Much of this material may be found in greater detail in D. L. DEXTER: *Proceedings of the Photoconductivity Conference at Atlantic City*, 1954 (New York, 1956), p. 155; H. Y. FAN, M. L. SHEPHERD and W. C. SPITZER: *Proceedings of the Photoconductivity Conference at Atlantic City*, 1954 (New York, 1956) p. 184; J. BARDEEN, F. BLATT and L. HALL: *Proceedings of the Photoconductivity Conference at Atlantic City*, 1954 (New York, 1956), p. 146.

[2] H. Y. FAN: *Phys. Rev.*, **82**, 900 (1951); H. M. JAMES: *Proceedings of the Photoconductivity Conference at Atlantic City*, 1954 (New York, 1956), p. 204.

[3] D. L. DEXTER: reference [1], p. 181 and 617.

[4] F. SEITZ: *Rev. Mod. Phys.*, **23**, 328 (1951); R. M. BLAKNEY and D. L. DEXTER: *Physical Society, « Defects in Crystalline Solids »*, Report of the 1954 Bristol Conference, p. 108; *Phys. Rev.*, **96**, 227 (1954).

[5] A. RADKOWSKY: *Phys. Rev.*, **73**, 749 (1948).

[6] J. BARDEEN *et. al.*: reference [1].

[7] G. G. MACFARLANE and V. ROBERTS: *Phys. Rev.*, **97**, 1714 (1955); **98**, 1865 (1955); W. C. DASH and R. NEWMANN: *Phys. Rev.*, **99**, 1151 (1955); E. BLOUNT, J. CALLOWAY, M. COHEN, W. DUMKE and J. PHILLIPS: *Phys. Rev.*, **101**, 563 (1956); H. Y. FAN *et al.*: reference [1].

[8] A. P. AIGRAIN: *Proceedings of the Amsterdam International Conference on Semiconductors*, in *Physica*, **20**, 978 (1954).

[9] F. J. DYSON: *Phys. Rev.*, **92**, 1331 (1953); H. SCHMIDT: *Phys. Rev.*, **105**, 425 (1957).

[10] F. URBACH: *Phys. Rev.*, **92**, 1324 (1953); F. MOSER and F. URBACH: *Phys. Rev.*, **102**, 1519 (1956); P. G. ALINE: *Phys. Rev.*, **105**, 406 (1957).

[11] D. B. DUTTON: to be published.

[12] S. TUTIHASI: *Phys. Rev.*, **105**, 882 (1957).

[13] For a more accurate description see D. L. DEXTER and W. R. HELLER: *Phys. Rev.*, **91**, 273 (1953).

[14] J. FRENKEL: *Phys. Rev.*, **37**, 17, 1276 (1931); *Physik. Zeits. Sowjetunion*, **9**, 158 (1936); R. PEIERLS: *Ann. der Phys.*, **13**, 905 (1932); J. FRANCK and E. TELLER: *Journ. Chem. Phys.*, **6**, 859 (1938).

[15] W. R. HELLER and A. MARCUS: *Phys. Rev.*, **84**, 809 (1951).

[16] F. SEITZ: *Modern Theory of Solids* (New York, 1940), pp. 415-417.

[17] *a*) D. S. MCCLURE and O. SCHNEPP: *Journ. Chem. Phys.*, **23**, 1575 (1955); *b*) J. W. SIDMAN: *Phys. Rev.*, **102**, 96 (1956).

[18] R. J. ZOLLWEG: *Phys. Rev.*, **97**, 288 (1955); for a review in English of much of the Russian and French work, see E. F. GROSS: *Suppl. Nuovo Cimento*, **3**, 672 (1956).

[19] T. MUTO and H. OKUNO: *Journ. Phys. Soc. Japan*, **11**, 633 (1956); T. MUTO: *International Congress on Theoretical Physics at Seattle*, September 1956. Further results will be published. A. W. OVERHAUSER: *Phys. Rev.*, **101**, 1702 (1956); P. L. HARTMAN, J. R. NELSON and J. G. SIEGFRIED: *Phys. Rev.*, **105**, 123 (1957); K. TEEGARDEN: *Phys. Rev.*, **108**, 660 (1957).

[20] A. S. DAVYDOV: *Journ. Exp. Theor. Phys USSR*, **18**, 210 (1948); H. WINSTON: *Journ. Chem. Phys.*, **19**, 156 (1951); D. FOX and O. SCHNEPP: *Journ. Chem. Phys.*, **23**, 767 (1955).

[21] W. van Sciver and R. Hofstadter: *Phys. Rev.*, **97**, 1181 (1956); K. Teegarden: *Phys. Rev.*, **105**, 1222 (1957).
[22] D. Fox and S. Yatsiv: *Phys. Rev.*, **108**, 938 (1957).
[23] D. Fox: to be published.
[24] *a*) Th. Foerster: *Ann. der Phys.*, **2**, 55 (1958); *b*) D. L. Dexter: *Journ., Chem. Phys.*, **21**, 836 (1953).
[25] T. Bodten: *Philips Research Rep.*, **6**, 425 (1951); **7**, 197 (1952).
[26] P. J. Leurgans: *Thesis*, University of Illinois (1952), unpublished.
[27] A. Smakula: *Zeits. f. Phys.*, **63**, 762 (1930).
[28] N. Inchauspé: *Phys. Rev.*, **106**, 898 (1957).
[29] L. Apker and E. Taft: *Phys. Rev.*, **79**, 964 (1950); **81**, 698 (1951); **82**, 814 (1951); **83**, 479 (1951); M. H. Hebb: *Phys. Rev.*, **82**, 701 (1951); D. L. Dexter and W. R. Heller: *Phys. Rev.*, **84**, 377 (1951); H. R. Phillipp and E. A. Taft: *Phys. Rev.*, **106**, 671 (1957).
[30] W. West and B. H. Carroll: *Journ. Chem. Phys.*, **19**, 417 (1951).
[31] M. Balkanski: *Bull. Am. Phys. Soc.*, **2**, 120 (1957).
[32] D. C. Reynolds, L. C. Greene, R. G. Wheeler and R. S. Hogan: *Bull. Am. Phys. Soc.*, **1**, 111 (1956).
[33] Several excellent reviews of various aspects of this subject have been given. See for example, M. Lax: *Proceedings of the Photoconductivity Conference at Atlantic City*, 1954 (New York, 1956); C. Herring: *Proceedings of the Photoconductivity Conference at Atlantic City*, 1954 (New York, 1956); C. C. Klick and J. H. Schulman: *Solid State Physics* (New York), Vol. **5**, p. 97 (1957).
[34] M. Lax and E. Burstein: *Phys. Rev.*, **100**, 592 (1955).
[35] K. Huang and A. Rhys: *Proc. Roy. Soc.* (*London*), A **204**, 406 (1950); S. I. Pekar: *Journ. Exp. Theor. Phys. USSR*, **20**, 510 (1950); **22**, 641 (1952); H. J. G. Meyer: *Physica*, **20**, 181, 1016 (1954); M. Lax: *Journ. Chem. Phys.*, **20**, 1752 (1952); R. C. O'Rourke: *Phys. Rev.*, **91**, 265 (1953).
[36] R. L. Bjork: *Phys. Rev.*, **105**, 456 (1957).
[37] A. von Hippel: *Zeits. f. Phys.*, 680 (1936).
[38] F. Seitz: *Trans. Faraday Soc.*, **35**, 79 (1939).
[39] D. L. Dexter: *Phys. Rev.*, **101**, 48 (1956).
[40] D. L. Dexter: *Phys. Rev.*, **96**, 615 (1954).
[41] I. S. Jacobs: *Phys. Rev.*, **93**, 993 (1954).
[42] R. W. Gurney amd N. F. Mott: *Trans. Faraday Soc.*, **35**, 69 (1939).
[43] D. L. Dexter, C. C. Klick and G. A. Russell: *Phys. Rev.*, **100**, 603 (1955).
[44] P. O. Löwdin: *Journ. Chem. Phys.*, **18**, 365 (1950).
[45] R. S. Knox and D. L. Dexter: *Phys. Rev.*, **104**, 1245 (1956).
[46] F. E. Williams: *Journ. Chem. Phys.*, **19**, 457 (1951).
[47] F. Seitz: *Journ. Chem. Phys.*, **6**, 150 (1938).
[48] D. A. Patterson and C. C. Klick: *Phys. Rev.*, **105**, 401 (1957).
[49] C. C. Klick: *Phys. Rev.*, **85**, 154 (1952) and subsequent publications.
[50] H. A. Jahn and E. Teller: *Proc. Roy. Soc.* (*London*), A **161**, 220 (1937); H. A. Jahn: *Proc. Roy. Soc.* (*London*), A **164**, 117 (1938).
[51] U. Öpik and M. H. L. Pryce: *Proc. Roy. Soc.* (*London*), A **238**, 425 (1957).
[52] J. von Neumann and E. Wigner: *Physik. Zeits.*, **30**, 467 (1929).
[53] R. V. Hesketh and E. E. Schneider: *Phys. Rev.*, **95**, 837 (1954); G. A. Russell and C. C. Klick: *Phys. Rev.*, **101**, 1473 (1956).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Metallic alloys.

J. FRIEDEL
Sorbonne - Paris

Introduction.

The metallic alloys are mixtures of elements, on an atomic scale, whose electrons are easily displaced or excited, as shown by their optical reflection, electronic conductivity and low temperature specific heat. Most of the elements are metallic; they mix rather freely, and form intermediary compounds in great number, so that the field of these lectures is enormous. There is however only little theoretical work done so far. As in every problem of physical chemistry, one can ask:

1) What is the *electronic* structure for a given nature and position of the nuclei?

2) What is the position of the *nuclei* in an alloy of given composition?

There are then the questions of the most stable crystallographic structure; of the exact equilibrium positions of the atoms (size effects); and of the vibrations of the atoms around these positions (vibrational entropy effects).

We will not consider the second type of questions, because they are mostly not specific of metallic alloys, and also because wave mechanical considerations have been little applied to them so far. We shall restrict ourselves to problems of electronic structure, and insist on the methods and approximations used. A systematic review of the results obtained would anyway be out of the question.

In part I, we recall the first order approximation methods developed mainly by MOTT and JONES [1] around 1935. In part. II, we study better approximations used when one wants more details concerning small perturbations introduced by alloying. In part. III, we study cases of large perturbations.

I.

First order approximations.

1. – Energies: the rigid band approximation.

1·1. *The model.* – When one is only interested in the energy spectrum of the alloy, the first order approximation may be called the rigid band approximation. It is based on the following assumptions:

a) Independent electrons, i.e. the conductive electrons are assumed to move all in the same potential, $V_L + V_p$ (V_L periodic potential, e.g. that of the pure matrix; V_p perturbing potential introduced by alloying).

b) First order perturbation: the change in energy by alloying for a Bloch function $\psi_K = u_K \exp[ik \cdot r]$ is then

$$\Delta E \cong \int_V \psi_k^* V_p \psi_k \, d\tau = \int_V u_k^* V_p u_k \, d\tau \, .$$

c) Nearly free electrons: if $u_K^* u_K$ is fairly constant in energy and in space,

$$\Delta E \cong \frac{1}{V} \int_V V_p \, d\tau = \text{const.}$$

This equation means that alloying shifts the energy of every state by a constant amount. This shifts by ΔE the whole curve of the density of states n versus energy E, without altering its shape. This is valid for electrons whose wave functions are near to those of free electrons; it requires V_p small and varying slowly within an electronic wave function. These conditions are roughly fulfilled by electrons near to the Fermi limit in usual metals.

1·2. *Transitional alloys* (MOTT, 1934). – This is one of the first applications of the model. When Ag is substituted to Pd, the supplementary electrons introduced shift the Fermi level towards the top of the incomplete d band (Fig. 1). They mainly fill the narrow d band, because of its higher density of states $n(E)$, thus reducing the strong paramagnetic susceptibility χ of Pd (Fig. 2). One needs about 50% Ag to suppress the paramagnetism, so that there are about 0.5 holes per atom in the d band of Pd [1]. The large temperature dependent part of the

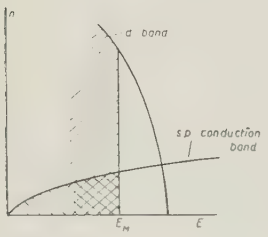


Fig. 1. – Band structure of Pd.

resistivity of Pd, due to phonon induced *s-d* transitions, is reduced in a similar way by alloying with Ag.

Such considerations are still widely used in transitional alloys. Thus in **Pd**U (*), χ varies as in **Pd**Ag but with a steeper slope, indicating that the six valence electrons of U fill the *d* band of Pd [2]. The susceptibility of Cr alloys, when plotted versus the average atomic number $\overline{N}$, shows a pronounced minimum near to pure Cr (Fig. 3), in agreement with the idea that the *d* band of Cr has a minimum in $n(E)$ near to the Fermi level [3].

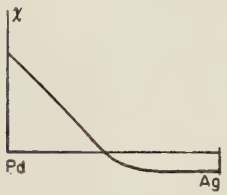


Fig. 2. – Magnetic susceptibility χ of PdAg alloys versus concentration *c*.

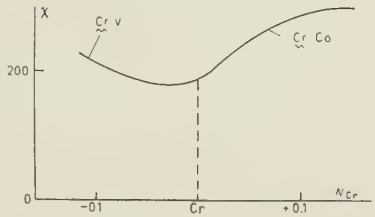


Fig. 3. – Magnetic susceptibility χ of Cr base alloys (**Cr**V and **Cr**Co) versus the average atomic number $\overline{N}$. χ in 10^6 emu/mole.

1·3. *Zone boundary effects* (Jones, 1935). – In *ordinary* metals such as Cu, where the conductive electrons have very nearly free electron wave functions, it may be shown that $n(E)$ only deviates from the parabola of free electrons, and the Fermi surface from a sphere, when this is near to a Brillouin zone boundary. The characteristic deviation which then occurs is pictured schematically in Fig. 4. When, by alloying, the Fermi limit reaches such a region, the average kinetic energy of the electrons in the alloy is somewhat smaller than for a gas of free electrons, or for other alloys with structures such that

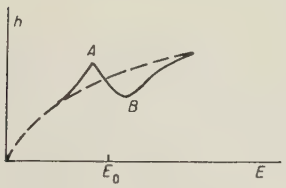


Fig. 4. – Deviations from the free electron parabola near a Brillouin zone boundary.

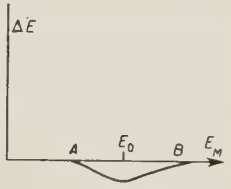


Fig. 5. – Stabilizing energy δE near a Brillouin zone boundary.

the hump *A* occurs for a larger electron per atom ratio. This small negative energy difference δE, pictured in Fig. 5, stabilizes the structure for which its maximum occurs near to the Fermi level E_M of the metal.

(*) **A**B means B dissolved in A.

These considerations allowed JONES to explain the « Hume Rothery rules » i.e. that in a great number of alloys the structures and boundaries in the phase diagrams depend mainly on the electron per atom ratio of the alloy [1]. Thus in CuZn brasses, the phase boundaries are near to the electron/atom ratios where a free electron Fermi sphere touches the Brillouin zone boundaries:

Phase	Electron/atom Inscribed sphere	Limit observed
α (F C C)	1.36	0 ÷ 1.38
β (B C C)	1.48	1.46 ÷ 1.50
γ (complex)	1.54	1.58 ÷ 1.66
ε C P H	—	1.78 ÷ 1.87
η C P H	—	1.94 ÷ 2

Such considerations are widely used by metallurgists, and are on the whole fairly successful in explaining complex binary and ternary diagrams, and complex structures such as CuZn γ (52 atoms per cell). Two restrictions must however be made:

a) These effects lead, in an exaggerated form, to the strong homopolar bindings of Ge, InSb, etc. But in ordinary alloys with small electron/atom ratios and close packed structures, they are very small, and only visible because the rest of the free energy varies very smoothly for these concentrations. Thus in CuZn alloys, the δE have been measured and are only of about 0.5 kcal/mole [4]. As a consequence, a phase like CuZn β is only stable at high temperature, and because of its large vibrational entropy. Therefore other effects (higher order corrections to the approximation used, size effects, etc.) might confuse the simple Jones picture.

The behaviour of lattice constants also shows that these are small effects. When the Fermi level E_M goes through a critical region E_0, one expects an anomaly of the type pictured in Fig. 6: the structure should deform so that E_0 comes nearer to E_M and more of the energy δE is gained. This leads to an anomalous expansion for $E_M < E_0$, to a concentration for $E_M > E_0$ [5]. A small anomaly of this type seems to occur for the *c* parameter in Mg alloys with small concentrations of elements of higher valencies such as Al, In [6]; it might be connected with an incipient overlap of the Fermi surface on the corresponding Brillouin zone boundary. On the other hand, no anomaly of the lattice parameter is observed for CuZn α alloys; one knows however from the anomalous skin effect that in Cu the Fermi surface is near to a Brillouin

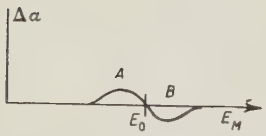


Fig. 6. – Corresponding anomaly expected for the lattice constant *a*.

zone boundary [7]; and the low temperature electronic specific heat of CuZn α alloys shows a peak around 10% Zn which suggests strongly that in Cu the Fermi level is just below the top of a hump such as A, Fig. 4 [8].

Similarly, in both Cu base and Mg base alloys, one observes no anomaly in the elastic constants or in the electrical properties [9, 10].

b) Contrary to the case of transitional alloys, there is here no direct relation between the *magnetic susceptibility* of an alloy and its density of states $n(E_M)$. This is because the diamagnetic terms are usually preponderant. Thus, in Cu, addition of Zn strongly increases the diamagnetism, although, for small concentrations, the Pauli paramagnetism probably increases, as the electronic specific heat does [11]. The main contribution here is probably an orbital diamagnetism, due to the fact that the conduction electrons are not distributed uniformly over all the atoms, but tend to concentrate on those of higher valency. To take such an effect into account, one has to study the spatial distribution of the electrons.

2. – Space: the screening effect.

When a solute atom is introduced into a metallic substance, the excess nuclear charge (*) Z displaces locally the mobile electrons until the displaced charge screens out the new nuclear charge exactly. We may say that the Z excess electron charges introduced with the solute atom remain around its nucleus, at short distances. What we call the excess charge Z is of course the charge of an interstitial impurity or the change of charge when an impurity is substituted in the matrix ($Z = 1$ for **Cu**H and **Cu**Zn). That this screening should be perfect follows, in classical physics, from the fact that no long range electric field can subsist in a conductor; in wave mechanics, it follows for instance from the fact that in a Coulomb field of potential $-Z/r$, there are an infinite number of bound states below the conduction band [12]; the conduction electrons would necessarily gain some energy by filling at least one of them, thus providing a localized screening.

As the screening should be perfect within the limits of classical physics, its radius should have atomic dimensions. A rough computation by MOTT [11], based on a Thomas-Fermi approximation, confirms this point.

2·1. *Thomas-Fermi screening* (MOTT, 1935). – This computation is necessarily based on a series of simplifications which will be discussed in the next part.

(*) Atomic units $e = m = \hbar = 1$ are used throughout.

1) Independent electrons (as before). All the electrons have the same potential $V_L + V_p$.

2) A Thomas-Fermi type of approximation. – If the perturbation V_p were a constant, the electronic density ϱ in the alloy would be related to the electronic density ϱ_0 in the pure matrix by

$$\varrho(E, \boldsymbol{r}) = \varrho_0(E - V_p, \boldsymbol{r}) . \tag{1}$$

Here $\varrho(E, \boldsymbol{r})$ refers to a density at a point $\boldsymbol{r}$ and *for the electrons up to energy* E. The approximation is to assume (1) to be still true when V_p varies in space. Thus, from Poisson's equation,

$$\Delta V_p = -4\pi[\varrho(E'_M, \boldsymbol{r}) - \varrho_0(E_M, \boldsymbol{r})] \simeq -4\pi[\varrho_0(E'_M - V_p, \boldsymbol{r}) - \varrho_0(E_M, r)] \tag{2}$$

if E_M and E'_M are the Fermi energies in the pure matrix and in the alloy respectively.

3) Linearization. For $|E'_M - E_M - V_p|$ small, we have

$$\Delta V_p \cong +4\pi \frac{\mathrm{d}\varrho_0}{\mathrm{d}E_M} (E_M + V_p - E'_M) . \tag{3}$$

4) Nearly free electrons. The electronic density $\mathrm{d}\varrho_0/\mathrm{d}E$ is then nearly a constant in space, and equal to its average value, the density of states $n(E)$. The solution of (3) with the right boundary conditions is

$$V_p = -\frac{Z}{r} \exp[-qr] , \tag{4}$$

with

$$\because E'_M = E_M, \; q^2 = 4\pi n(E_M) , \tag{5}$$

and a screening charge density

$$\varrho(E_M, r) - \varrho_0({}_M E, r) = -\frac{1}{4\pi} \Delta V_p = \frac{Zq^2}{4\pi} \frac{\exp[-qr]}{r} . \tag{6}$$

2·2. *Qualitative consequences.* – The preceeding formulae are an extension to Fermi gases of the classical Debye Hückel model of electrolytes. They provide a rough but very useful estimate of the screening radius.

2·2.1. Localization. – Due to screening, the Coulomb field of the nuclear charge Z is cut-off at a distance of the order of a few times $1/q$. For most metals, this is of the order of the interatomic distance. Thus atoms

of the matrix which are not neighbours to impurity atoms do not feel their presence, except for size effects not discussed here. Experiments which measure the properties of the bulk of the matrix are unaffected by the presence of impurities.

Thus:

a) The optical and X-ray absorption edges of Cu are unaffected by the admixture of solute atoms such as Zn up to several percent (constancy of E_M) [13, 37].

b) The Knight shift of Cu in **Cu**Al seems to remain a constant up to 10% of Al (constancy of $\mathrm{d}\varrho/\mathrm{d}E_M$ on the copper atoms) [14, 15]. Recent measurements on **Ag**Cd [39] indicate however some shift with Cd concentration of both the Ag and Cd Knight shifts. But this might be due to a size effect not discussed here.

2·2.2. Size of the screening. – If r_a is the radius of the atomic sphere ($4\pi r_a^3/3 =$ the atomic volume), the part of the screening outside the atomic sphere of the impurity is, from (6),

$$Z_{\mathrm{ext}} = \int_{r_a}^{\infty} 4\pi r^2 \cdot \frac{Zq^2}{4\pi} \frac{\exp[-qr]}{r} \mathrm{d}r = Z(1 + qr_a) \exp[-qr_a] . \tag{7}$$

Three cases may be distinguished:

a) In *transitional* metals, $n(E_M)$, thus q, are large, so that Z_{ext} is vanishingly small. The screening occurs on the impurity atom itself. Thus two impurity atoms do not feel the presence of each other, even when nearest neighbours. This complete *lack of interaction* between solute atoms leads to « *ideal* » solid solutions (no energy of dissolution and perfect disorder). This is indeed a characteristic of such alloys, when size effects are negligible (e.g. NiCo alloys).

b) In *ordinary* alloys, like CuZn, $q \simeq 1$ AU and $r_a = 3$ AU, thus $Z_{\mathrm{ext}} \simeq$ $\simeq$ a few $(1/10)Z$: the solute atoms with $Z > 0$ give up a small part of their excess electrons to the neighbouring atoms in the matrix. They are a little more « electronegative ».

This overlap leads to a small *interaction* between solute atoms: the perturbation V_p when two solute atoms are present is the sum of those for the isolated solute atoms, because equation (3) is linear in V_p; and the computation, within first order approximation, of the potential and kinetic energies of the system leads to an energy of interaction

$$\Delta U = \frac{Z_1 Z_2 \exp[-qd]}{d} , \tag{8}$$

if the two impurities, with excess charges Z_1 and Z_2, are a distance d apart [17]. This leads to an attraction when Z_1 and Z_2 are of opposite signs and to a repulsion for Z_1 and Z_2 of the same sign (*). Such an interaction should lead to a small short range order in the solid solution. It has been used by LAZARUS to explain why, in Ag, polyvalent impurities ($Z > 0$) diffuse with a smaller activation energy than pure Ag, while transitional impurities like Pd have a larger activation energy ($Z < 0$). He assumes that a vacancy in Ag acts as an impurity with $Z = -1$, thus as Pd, and therefore is attracted by polyvalent impurities and repelled by Pd [18]. Formula (8) can however be only very rough, because it depends sensitively on the screening radius.

Finally in a high concentration alloy like CuZn β (50 % Zn), each Zn atom has a net charge $+Z_{\text{ext}}$ and each Cu atom a charge $-Z_{\text{ext}}$. The alloy should then be more stable in an *ordered* state; the stabilizing Madelung energy is in this case equal to $-10\,Z_{\text{ext}}^2$ eV. The alloy is indeed ordered in the low temperature range, with a stabilizing energy, of about 0.1 eV. As pointed out by MOTT [1], this shows that the charge (Z_{ext}) in each atom is small, of the order of $(1/10)Z$.

c) In *some polyvalent* metals (e.g. graphite, bismuth), the density of states $n(E_M)$ is anomalously low, because E_M falls at a strong minimum such as B, Fig. 4. One then expects large screening radii and fairly strong interactions between solute elements. No observations have been made so far on this point.

2 3. *Ferromagnetic alloys. Equivalence with the rigid band model.* – According to MOTT, ferromagnetic Ni has a band structure analogous to Pd, with one half of the 3*d* band, with say spin down, overlapping the Fermi level E_M and the other half, with spin up, full, and thus below E_M (Fig. 7).

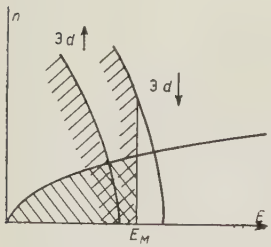


Fig. 7. – Band structure of ferromagnatic Ni and Ni base alloys.

When an impurity is dissolved into it, its excess nuclear charge Z is therefore screened out within the impurity atom itself, the screening being provided by the part of the d band with large $n(E_M)$, thus spin down. By alloying, moments of the matrix and of the impurity should then be *independent of concentration*:

$$\begin{cases} \mu_{\text{matrix}} = \text{const}\,. \\ \mu_{\text{impurity}} = \mu_{\text{matrix}} - Z\mu_{\text{B}}\,, \end{cases} \tag{9}$$

(*) This formula was first proposed by LAZARUS [18], by an incorrect argument which left out the kinetic and part of the potential energies. Another formula, proposed previously by the author [19], is also incorrect because it leaves out an interaction between the screening electrons and the nuclei.

(μ_B Bohr magneton). In **Ni**Fe alloys for instance, $Z = -2$, thus $\mu_{Ni} = 0.6\mu_B$ and $\mu_{Fe} = 2.6\mu_B$. This is indeed observed by neutron diffraction experiments in the disordered solid solution up to 50 % Fe, and also on ordered Ni_3Fe where results are more accurate (Fig. 8 and 9 [20]).

From (9), the *average* magnetic moment of an alloy with atomic concentration c should be

$$\bar{\mu} = \mu_{\text{matrix}} - Zc\mu_B \,. \tag{10}$$

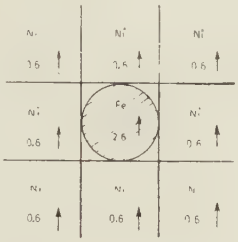


Fig. 8. – Spatial distribution of the magnetic moments in ferromagnetic **Ni**Fe.

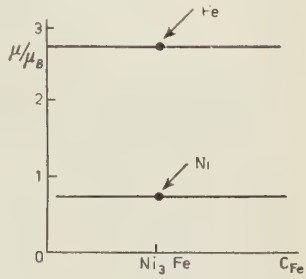


Fig. 9. – Magnetic moments on Ni and on Fe in **Ni**Fe, according to neutron diffraction.

This is in good agreement with the straight line AB with 45° slope obtained when $\bar{\mu}$ is plotted versus the average atomic number in various Ni and Co base alloys (Fig. 10). Now equation (10) was obtained by MOTT [1] just using the rigid band model. This shows that the screening considerations are not contradictory to the rigid band model. It is actually easy to show that the equations (1) to (5) for screening lead to the rigid band model (*). The two models developped so far are therefore complementary: they are the energetic and the spatial aspects of the *same* first order approximation.

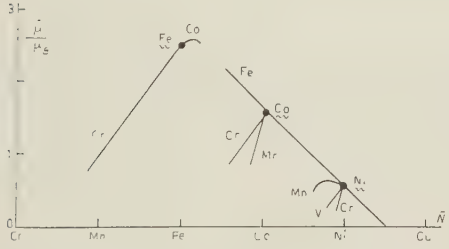


Fig. 10. – Average magnetic moments $\bar{\mu}$ versus average atomic number $\bar{N}$ in ferromagnetic alloys.

(*) Within the approximations of Sect. 2·1, the density of states n with energy E is such that

$$\int^{E} n(E)\,\mathrm{d}E = \frac{1}{V}\int_V \varrho(E, \boldsymbol{r})\,\mathrm{d}\tau = \int_V \varrho_0(E - V_p\,\boldsymbol{r})\,\mathrm{d}\tau = \int_V \varrho_0(E+\Delta E, \boldsymbol{r})\,\mathrm{d}\tau -$$

$$-\int (\Delta E + V_p)\,\frac{\mathrm{d}\varrho_0}{\mathrm{d}E}\,\mathrm{d}\tau = \int^{E} n_0(E+\Delta E)\,\mathrm{d}\tau \qquad \text{if} \qquad \Delta E = -\frac{1}{V}\int_V V_p\,\mathrm{d}\tau \,.$$

n is the density of states in the alloy, n_0 that in the pure matrix.

II.

Better approximations for small perturbations.

The methods developed above give a picture of the alloys very useful in many ways. It allows us to study homopolar binding effects (Hume-Rothery alloys), ionic effects (Mott's theory of order), short range order effects, it shows them to be small, especially in transitional alloys; and it relates these properties to the density of states at the Fermi level $n(E_M)$, a quantity of evidently primary importance in a conductor.

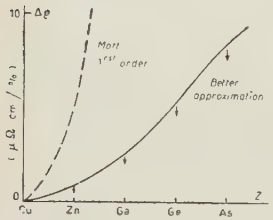


Fig. 11. – Residual resistivities of impurities in copper. The crosses are the experimental values.

The methods are however insufficient in a number of problems, which require a more accurate knowledge of the electronic structure. One can for instance compute the *electrical resistivity* $\Delta\varrho$ of the conduction electrons of a metal, due to the presence of impurities with the perturbation potential V_p obtained above. Within first order approximation, $\Delta\varrho$ should be computed with the Born approximation [1]. One obtains then for Cu alloys a resistivity about 10 times larger than the observed one (Fig. 11). Other discrepancies of this kind will be presented below. They give a physical interest to a systematic discussion of the approximations made in Part I. This will be done presently.

1. – Independent electrons.

A systematical study of correlation corrections, along the lines described by PINES in his lectures, is still lacking. Within first order approximation methods, it is reasonable to introduce a Thomas-Fermi type of correction, both for exchange and Coulomb correlations [21]. This would lead to the following effects:

a) A relatively *small* correction on the *screening radius*, making it smaller for $Z > 0$ and larger for $Z < 0$ (because correlation effects make each electron feel the screening somewhat less) ([19].

b) *Large* corrections on the screening *energy*, and on the *magnetic* properties (as in pure metals). This certainly explains why all efforts to compute energies of solution in this way have failed so far [22].

2. – Linearization.

It is easy to introduce terms of higher orders in the development of the screening charge. These take into account not only $n(E_M)$, but dn/dE_M, d^2n/dE_M^2, etc. These terms are certainly not negligible even for « small » perturbations V_p e.g. for $Z = 1$ (**Cu**Zn). The following effects are obtained [19]:

a) All the bands tend to broaden and to have smaller densities of states $n(E)$.

b) For « electrons » ($dn/dE_M > 0$), the screening radius becomes smaller for $Z > 0$, larger for $Z < 0$; the reverse is true for « positive holes » ($dn/dE_M < 0$).

3. – Thomas Fermi approximation.

This is probably the most interesting point to consider. The approximation is based on a mathematical theorem, due to VON LAUE [23] which we analyze first.

3·1. *Von Laue's theorem* (1914). – This refers to a gas of *free* electrons enclosed in a box of surface S. The theorem states that the *electronic density* $\varrho(E, \boldsymbol{r})$ (density at $\boldsymbol{r}$ of electrons with energy less than E) *is independent of the boundary conditions on* S *at all points* $\boldsymbol{r}$ *in the box which are not too near to the surface* S.

This is proved by analyzing ϱ in *orthogonal packets* of the electronic wave functions ψ with energy less than E. The size of the packets, thus their spatial density, are evidently independent of the presence of S and of the boundary conditions on them, as long as their distance to S is much larger than their size: von Laue's theorem does *not* apply to a boundary layer along the surface S. *The thickness of this layer is given by the size of the wave packets*; *it is thus of the order of the wave length of the electrons of energy* E.

As pointed out by VON LAUE himself, this theorem is applicable to any wave mechanical problem, and therefore to phonons, electrons in a periodic potential, metallic alloys....

3·2. *Examples.* – Two simple examples might help to understand this result.

3·2.1. One dimensional square box. – In this case, a direct computation of the electronic density $\varrho(E, r)$ shows that, for impenetrable walls ($\psi = 0$ on S), ϱ starts from zero at the wall and takes a very nearly constant value at distances from the wall larger than $\pi/2k$ ($k^2 = 2E$). This constant value ϱ_0 is that obtained for Born-von Karman conditions (Fig. 12). The

deviation from this constant value is plotted in Fig. 13. It is given by

$$\frac{\varrho}{\varrho_0} - 1 = \frac{\sin 2kx}{2kx} .$$

This shows that the interference effects obtained for $x > \pi/2k$ are very small. A repulsion of this kind is the major part of the *surface energy* of metals.

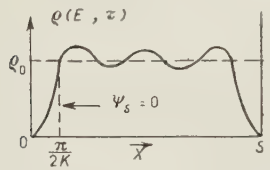


Fig. 12. – Electronic density $\varrho(E, x)$ in a box with impenetrable walls.

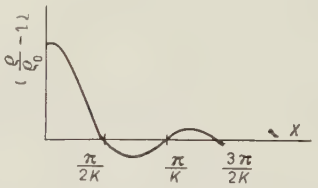


Fig. 13. – Relative deviation from a constant value for the electronic density near to an impenetrable wall.

3 2.2. Thin spherical well. – This situation is pictured in Fig. 14, against the radial distance r to the center of the well. If ϱ is the electronic density, ϱ_0 that if there were no well ($V_0 = 0$), the electronic density $\varrho - \varrho_0$ displaced in the gas by the presence of the well is given by

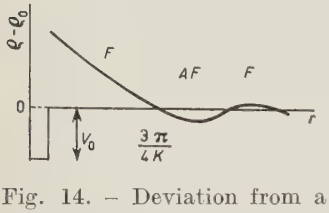


Fig. 14. – Deviation from a constant value for the electron density around a thin spherical well (only the part outside the well is pictured).

$$(11) \quad \varrho - \varrho_0 \cong \frac{4U_0k^4}{\pi^3} \frac{\sin 2kr - 2kr\cos 2kr}{(2kr)^4} ,$$

if

$$k^2 = 2E \qquad \text{and} \qquad U_0 = \frac{4\pi}{3} a^3 V_0 .$$

This approximation formula is only valid for ka and $\sqrt{2V_0}\cdot a \ll \pi$. Here again, the interference fringes are very week, except for a central spot of radius $\simeq \frac{3}{4}\pi/k$.

Formula (11) is at the origin of what may be called the « *Fröhlich-Bloembergen-Kittel* » *effect* [24]: the contact part of the magnetic interaction of a nucleus with the conductive electrons may probably be represented by such a potential well V_0 acting on one half of the electrons, with spins say upwards, and a similar potential $-V_0$ of opposite sign acting on the other spin direction. Then, within first order approximations, there is *no* excess electronic charge around the nucleus, but an excess of electrons with spin upwards near to the nucleus, given by *twice* the expression (11). This « spin polarization »

will induce a tendency to *ferromagnetic* alignement on a neighbouring nucleus, if it is at a distance r such that $\varrho - \varrho_0 > 0$ (F regions, Fig. 14); an antiferromagnetic one otherwise (AF regions). At the Fermi level of usual metals, $3\pi/4k$ is slightly larger than the internuclear distance. There should be a slight tendency to ferromagnetic alignement of the nuclei. This seems actually to exist and to affect the width of their resonance level [25]. The effect seems much stronger in polyvalent metals like Tl where the conduction electrons deviate markedly from free electrons; in such metals, the « effective » k may be smaller, because it should be counted from a band edge higher than the bottom of the valence band.

3 3. *Application to impurity atoms.* – The preceding examples show clearly that corrections to the Thomas-Fermi approximation should spread the screening charge somewhat. This effect should be *small* except: *a*) near to the nucleus, where V_p varies rapidly with r; *b*) and also at large distances, if the Thomas-Fermi screening radius is much smaller than $3\pi/4k_M$, where k_M is the wave number for Fermi electrons.

For a given V_p, these effects can be studied by *treating the conduction electrons as free and only scattered by* V_p. One then gets a useful formula, which we shall first recall.

3 3.1. The phase shifts sum rule. – When scattered by a spherically symmetrical potential $V_p(r)$, free electrons have wave functions which may be analyzed in spherical harmonics [26]. The radial part $\varphi_l(r)$ of the l-th component with energy E satisfies the equation:

$$\frac{\mathrm{d}^2\varphi_l}{\mathrm{d}r^2} + \left[2(E - V_p) - \frac{l(l+1)}{r^2}\right]\varphi_l = 0 .$$

A combination of this equation with that for another wave function φ_l' of different energy E' leads to

$$2(E' - E)\int_0^R \varphi_l \varphi_l' \,\mathrm{d}r = \int_0^R \left(\varphi_l' \frac{\mathrm{d}^2\varphi}{\mathrm{d}r^2} - \varphi_l \frac{\mathrm{d}^2\varphi_l'}{\mathrm{d}r^2}\right)\mathrm{d}r = \left[\varphi_l' \frac{\mathrm{d}\varphi_l}{\mathrm{d}r} - \varphi_l \frac{\mathrm{d}\varphi_l'}{\mathrm{d}r}\right]_0^R .$$

Setting $2E = k^2$ and for $E' \to E$, this gives

$$2k\int_0^R \varphi_l^2 \,\mathrm{d}r = \left[\frac{\mathrm{d}\varphi_l}{\mathrm{d}r}\frac{\mathrm{d}\varphi_l}{\mathrm{d}k} - \varphi_l \frac{\mathrm{d}^2\varphi_l}{\mathrm{d}k\,\mathrm{d}r}\right]_0^R .$$

Using the fact that $\varphi_l \to Ar^{l+1}$ for $r \to 0$ and $\varphi_l \to B \sin\,(kr + \eta_l(k) - \frac{1}{2}l\pi)$

for $r \to \infty$, one has

$$\int_0^R \varphi_l^2 \, \mathrm{d}r \to \frac{B}{2k}\left\{k\left|R + \frac{\mathrm{d}\eta_l}{\mathrm{d}k}\right| - \frac{1}{2}\sin 2\left(kR + \eta_l - \frac{l\pi}{2}\right)\right\} \qquad \text{for } R \to \infty.$$

η_l is « the phase shift » produced on φ_l by the action of the perturbation V_p [26]; A and B two normalizing constants.

When $V_p = 0$, the corresponding equation for the unperturbed spherical harmonic φ_{0l} with the same energy is:

$$\int_0^R \varphi_{0l}^2 \, \mathrm{d}r \to \frac{B_0}{2k}\left\{kR - \frac{1}{2}\sin\left(2kR - l\pi\right)\right\} \qquad \text{for } R \to \infty.$$

The screening charge produced per unit k in the sphere of radius R will be obtained by the difference of these two expressions multiplied by the corresponding densities of states per unit k. For φ_{0l}, it is easily seen that the density is $4R/\pi B_0$. From von Laue's theorem, the density of states for φ_l must be $4R/\pi B$, so as to give the same asymptotic value for large R's. One finds then *for the screening charge per unit k in a sphere of large radius R* the following expression [13]:

$$(12) \qquad \int_0^R \frac{\mathrm{d}(\varrho - \varrho_0)}{\mathrm{d}k} 4\pi r^2 \, \mathrm{d}r = \frac{2}{\pi}\sum_l (2l+1)\left[\frac{\mathrm{d}\eta_l}{\mathrm{d}k} - \frac{1}{k}\sin\varphi_l \cos\left(2kR + \eta_l - l\pi\right)\right].$$

Account has been taken here of the degeneracy of the l-th spherical component; one has summed over all values of l; and assumed two electrons (with opposite spin directions) per state.

3·3.2. The oscillating term. – The preceeding formula shows that the screening charge around a perturbation is the sum of a constant term and of a term which varies sinusoidally with R. By summing this *second* term over k and differentiating with respect to R, for a thin spherical well, one sees easily that this leads to the interference pattern of Fig. 14, discussed in Sect. 3·2.2. [27].

More generally, for any given perturbation $V_p(r)$, the oscillating term in equation (12) will give rise to similar interference fringes, which should be computed in the same way. This has not been done so far, but the resulting effect should *not be very important*:

a) In the case of Fig. 14, only η_0 differs from 0. For more extended V_p, other phase shifts will be significant, and swamp somewhat the long range oscillations, which therefore will be even less marked.

b) Only *close range* effects ($r < 3\pi/4k_M$) can therefore be significant. These can obviously be important only when the Thomas-Fermi screening radius of Part. I is much shorter than this distance. Numerical computations show that this is not the case in usual metals such as copper where these effects are negligible, at least for « small » perturbations (cf. for instance [13]). However, as the Thomas-Fermi screening varies as $1/q$, thus as $(m^*k_M)^{-1}$ for electrons with an effective mass m^*, one expects the actual screening to be much more spread out than the Thomas-Fermi one in transitional metals (*) (m^* large) or in « semi metals » such as Bi (k_M very small) [36].

Another possible case when the oscillating terms might spread the screening significantly is in the resonance effects discussed below in Part. III.

The *Zener effect* is a coupling of neighbouring magnetic ions through the spin polarization they induce in the conduction band [28], by the same mechanism as in the nuclear Fröhlich effect discussed above. If it exists, it is therefore likely to be small.

3 3.3. The screening condition. – As the oscillating term in (12) has an average value zero, it is the term in $\mathrm{d}\eta_l/\mathrm{d}k$ which gives the charge displaced by V_p in the gaz.

An integration over k gives, for *the total charge locally displaced by* V_p *below energy* E:

$$\int_0^\infty [\varrho(E, \boldsymbol{r}) - \varrho_0(E, \boldsymbol{r})] 4\pi r^2 \,\mathrm{d}r = \frac{2}{\pi} \sum_l (2l+1)\eta_l(E) \,. \tag{13}$$

This is a very useful rule when one looks for a potential V_p better than the Thomas Fermi one. As a first step towards selfconsistency, one can choose V_p *such that the total screening charge below the Fermi level is equal to* Z *electronic charges:*

$$\frac{2}{\pi} \sum_l (2l+1)\eta_l(E_M) = Z \,. \tag{14}$$

3 3.4. Electrical resistivity of impurities. – Using then any likely approximate potential V_p which satisfies (14), and computing the corresponding residual resistivity by the exact equation

$$\Delta\varrho = \frac{4\pi c}{p k_M} \sum_l l \sin^2 (\eta_{l-1} - \eta_l)_{E_M} \,, \tag{15}$$

(p electrons per atom in the matrix, c atomic concentration of impurities), one obtains a much better agreement with experiment for alloys such as

(*) In transitional metals, the neutron diffraction data discussed in Part. I indicate that most of the screening still occurs within the impurity atoms.

Zn, Ga, Ge, As in Cu (Fig. 11). The remaining discrepancies are probably to be attributed mainly to the factors neglected (corrections for size effects, effective mass, etc.) [29].

3 3.5. Electronic density near to the nucleus. – The Thomas-Fermi approximation of Part I leads to an electronic screening density which is infinite at the origin. This is clearly wrong from the Knight shift experiment on impurity atoms (Al in **Cu**Al [14]) and difficult to reconcile with other measurements such as positron annihilation or X-ray spectra in metals. A more exact computation of the scattered wave functions is needed in this case (cf. [38] for positron annihilation); it leads to a *finite* density at the origin and reasonable agreement with experiment.

3 4. *Nearly free electrons.* – The conduction electrons have been treated so far as free or nearly free. This is a reasonable assumption for ordinary metals. One expects therefore only small corrections when a fuller account is taken of the periodic potential of the matrix. A generalization of the concept of phase shifts has thus been carried out recently in an approximate manner [30]. It gives results not very different from the free electron values, and still definitely larger than the experimental ones, as the following table shows (in $\mu\Omega\,\mathrm{cm}/\%$):

	$\Delta\varrho_{\mathrm{th}}$	$\Delta\varrho_{\mathrm{exp}}$
Cu Zn	0.58	0.34
Ag Cd	0.56	0.38
Au Hg	0.62	0.41

From a general point of view, writing the residual resistivity as

$$\Delta\varrho = (N_{\mathrm{eff.}}\tau)^{-1}\,,$$

one expects the usual effective mass m^* of the conductive electrons to enter in their effective number $N_{\mathrm{eff.}}$, while an effective mass nearer to the normal mass m should enter in the relaxation time τ, because the perturbation V_p is so localized [29].

III.

Strong perturbations.

1. – The period factor.

The simplest case of a strong perturbation is in an alloy such as **Ag**Au, where the solute element has the same valency as the matrix, but belongs to a different period.

One expects in such a case the two elements to mix up freely, without altering very much their electronic structure: the wave functions at the bottom of the conduction band of Ag and Au should be easily connected, because they have nearly the same energy and the same amplitude on the boundaries of the atomic cells the big perturbation potential energy introduced with Au is nearly compensated by a big change in kinetic energy due to the supplementary node around the Au nucleus. This node occurs of course because of the supplementary 5*s*, 5*p*, 5*d* shells of the Au atom (Fig. 15). Such alloys have indeed *small energies of formation* and *very small electrical resistivities*. These have been computed treating the conductive electrons as free [1, 31] and, more recently, taking the periodic potential of the matrix into account in an approximate fashion. The following table shows that the agreement is good in alloys such as AgAu when no other factors such as differences in size or valency play a role [30] ($\Delta\varrho$ in $\mu\Omega\,\mathrm{cm}/\%$).

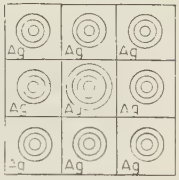


Fig. 15. – Nodes of the wave function at the bottom of the conduction band in **Ag**Au alloy.

	$\Delta\varrho_{th}$	$\Delta\varrho_{exp}$
Ag Au	0.46	0.38
Au Ag	0.41	0.38

2. – Transitional impurities.

Alloys with transitional impurities often exhibit peculiar properties which are probably related to the presence of *d* bound states or virtual bound states. The concept of virtual bound states is well known in the scattering of free electrons by free atoms of a gaz [26, 32]. This is first recalled.

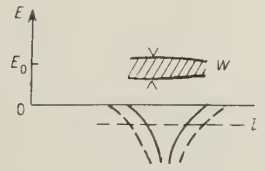


Fig. 16. Real bound state dashed curves) and virtual bound state (continuous curves) in an energy versus space diagram.

2·1. *Virtual bound states and resonance scattering for free electrons scattered by free atoms.* – Let us consider an atom with a spherically symmetrical potential not quite strong enough to accept a certain bound state with quantum number *l*. That is, if it was slightly more attractive, as pictured by the dashed curve in Fig. 16, it would accept an *l*-th bound state. When the potential is reduced from the value of the dashed curve to the actual value cor

responding to the continuous curve, the bound state increases in energy and merges into the continuum of extended states. It is often useful to think of it as still existing, with a positive energy. But as it has now the same energy as an extended state, it will resonate with its l-th spherical component to build up two (extended) states of slightly different energies; these in turn will have the same energies as extented states, with whom they will resonate etc. Finally this « virtual » bound level broadens into *a region in energy and space* corresponding to something like the shaded region in Fig. 16, where:

a) The amplitude of the l-th spherical component of the extended states is anomalously large.

b) These large amplitudes correspond to an excess of charge density which, summed over the whole shaded region, is approximately equal to that of the initial l-th bound state (with degeneracy $2l+1$).

This second point is an obvious consequence of the resonance mechanism. It may also be seen by considering the behaviour of the *phase shift* η_l of the l-th component of the extended states. For the punctuated potential of Fig. 16, η_l must start from π for $E = 0$, because there is a bound state [32], as pictured in Fig. 17. For the continuous potential without bound state, η_l starts from 0 for $E = 0$; it must join the dashed curve for $E \gg E_0$, because there it is surely little sensitive to a small change in the potential. η_l should therefore vary rather rapidly from nearly 0 to nearly π in a region of width w near to energy E_0. Numerical computations confirm this point [26]. From the screening r lation of Part II, one sees that such a variation of η_l corresponds to increasing by 1 state l (with $2l+1$ degeneracy) the screening charge in the energy range of the virtual bound level.

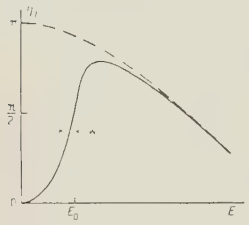


Fig. 17. – Phase shift $\eta_l(E)$ for a real bound state (dashed curve) and for a virtual one (continuous curve).

Because of the strong resonance, electrons with energy near to that E_0 of the virtual bound state are strongly scattered. This is what makes the interest of these levels for free atoms. The relation given in Part II for $\Delta\varrho$ shows that the *residual resistivity* of a Fermi gas with Fermi energy E should indeed go through a *peak* when E is equal to E_0, Fig. 18: η_l crosses the value $\pi/2$ at an energy E_0

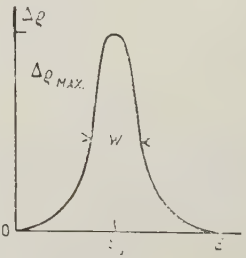


Fig. 18. – Resonance scattering peak corresponding to a virtual bound level of average energy E_0 and width w.

where all the other phase shifts are small or near to multiples of π (if E_0 is small). Neglecting completely the contribution of the other phase shifts, the relation of Part II gives for the maximum resistivity, with the same notations,

$$\Delta\varrho_{max} = \frac{4\pi c}{p k_M} (2l + 1) . \tag{16}$$

This peak is large compared with the resistivities for the « usual » alloys discussed in Part II, where $\sum_l l \sin^2 (\eta_{l-1} - \eta_l) \ll 2l+1$.

Finally the *width* w of the level increases obviously with the amount of l-th spherical component in the free electron state of energy E_0.

Thus:

a) At a given energy E_0, *w decreases for increasing* l; for $l = 0$ it is actually so large that virtual s levels have no physical significance.

b) For a given l, it is *roughly proportional to* E_0 (for small values of E_0); for $l = 2$ (d states), it is of the order of 2 eV for $E_0 = 7$ eV (i.e. the Fermi energy of copper).

We shall now discuss some applications of these concepts to analogous cases in alloys.

2·2. *Transitional impurities in aluminium.* – When an element such as Ti, on the left of a transitional series, is dissolved in aluminium, we expect its d shell to be unstable, thus to be practically empty and above the Fermi level of aluminium. For an element such as Ni or Cu, the d shell should be stable, full and below the Fermi level of aluminium. Thus, when going continuously from Ti to Ni, the states corresponding to the d shell of the transitional impurity should cross the Fermi level of aluminium, and give rise to one or several « virtual » bound d states, by mixing with the conductive states of aluminium.

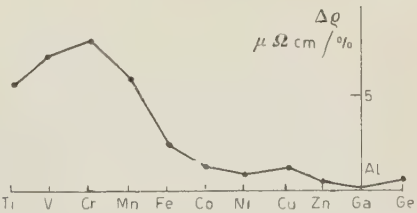


Fig. 19. – Residual resistivities of transitional impurities in aluminium.

The residual resistivities $\Delta\varrho$ of these impurities in aluminium are plotted in Fig. 19. They show indeed a large and broad resistivity peak around Cr, which we interprete as due to resonance scattering of the Fermi electrons of Al by the broadened d shell of the transitional impurity when this crosses the Fermi level of Al. The top of the peak reaches a value near to that predicted by formula (16) for $l = 2$, $p = 3$ and k_M that of Al.

The resistivity peak is very broad and continuous, indicating that, in this case, each d shell gives rise to only *one* virtual bound d level, with room for 10 electrons. One could a priori have expected this level to split into a succession of levels of different energies, from a variety of reasons:

a) The cubic symmetry of the lattice should split the xy type of d states from the $x^2 - y^2$ one.

b) The Coulomb correlation between d electrons of the same shell should stabilize the intermediary states of ionization $3d^9$, $3d^8$, $3d^7$, ..., $3d^2$, $3d$ [33].

c) The exchange correlations within a shell should induce one half of the d shell, with one spin direction, to become empty before the other half (Hund's rule).

All these splittings are apparently swamped here by the *resonance broadening*. This is *very large* when the d state crosses the Fermi level, because the Fermi energy of Al is large ($\simeq 12$ eV, leading to d states with a width $w \simeq 4$ eV if the conductive electrons are treated as free).

In conclusion, we picture here the d shell of the transitional impurity as one broad virtual level, with room for 10 electrons of both spin directions, which empties itself continuously when one goes from Ti to Ni (Fig. 20). The number of d electrons or d holes for a given element is not necessarily an integer. The presence of this virtual level across the Fermi level E_M gives rise both to a large Pauli paramagnetism (due to the large increase in the density of states at E_M) and to a large orbital diamagnetism (due to the d shells). The two contributions are *independent of temperature*, as is indeed observed for the *magnetic susceptibility* of all these alloys [29].

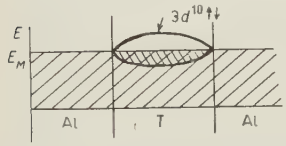


Fig. 20. – d shell of a transitional impurity T in aluminium, broadened by resonance with the conductive states.

2·3. *Transitional impurities in* Cu, Ag *or* Au. – As pointed out above, exchange interactions within a d shell tend to split it into two halves, of opposite spin directions, which would empty themselves one after the other when going from Ni to Ti. This splitting can only occur if the energy of splitting is larger than the width w of the states. The energy gained per d electron from the splitting is $p\,\Delta E$, if p is the number of electrons (if less than 5) or holes (otherwise) in the d shell; ΔE is the average energy gained when two d electrons with antiparallel spins are put with their spins parallel. The condition for Hund's rule to apply is thus

$$w \leqslant p\,\Delta E\,. \tag{17}$$

From atomic spectra, ΔE is 0.6 to 0.7 eV. As $p \leqslant 5$, the exchange splitting is impossible in aluminium, where $w \simeq 4$ eV according (*) to Sect. 2·1. It should however occur in matrices with a smaller Fermi energy, where w is smaller. This would lead to a double peak of $\Delta\varrho$ in the series from Ti to Ni, which is indeed observed for transitional impurities in noble metals (Cu, Ag or Au), as shown in Fig. 21: peak A would correspond to the emptying of the first half A of the d shell, peak B to the second half B with opposite spin direction, as pictured in Fig. 22. Here again, the top of the resistivity peaks are near to what is predicted by equation (16) (+).

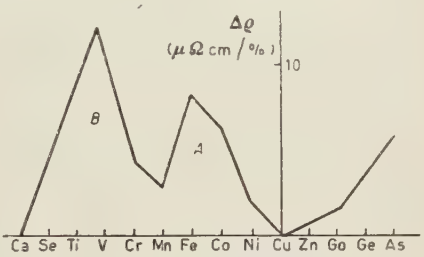


Fig. 21. – Residual resistivities of transitional impurities in copper.

For Cu, the width w of the d states should be of the order of 2 eV. One expects then from condition (2) the splitting of Fig. 22 to occur for $p \geqslant w/\Delta E = 3$; thus for impurities like Fe, Mn, Cr, but *not* for Ni and Co, where the number p of d holes is too small.

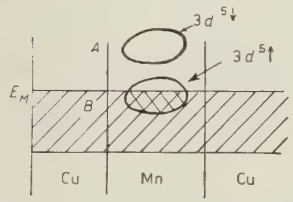


Fig. 22. – d shell of a transitional impurity such as Mn in Cu, split into two virtual bound levels of opposite spin directions.

This is in agreement with the *magnetic* properties:

a) For Fe, Mn, Cr, the magnetic susceptibility χ varies strongly with temperature, following a *Curie-Weiss law* $\chi = \text{const}/(T - \text{const})$, as expected for a split d shell such as in Fig. 22. For Mn, between the two resistivity peaks, Fig. 21, the Curie constant corresponds indeed to a magneton number near to (×) 5.

b) For Ni and Co (at least at low concentrations), the magnetic susceptibility is *independent of temperature*, as in the Al alloys pictured in Fig. 20 [34].

(*) The width w should be about twice as large in Al as in Cu, because the Fermi energy is 13 eV instead of 7 eV.

(+) A factor $(p+1)/2p$ must be introduced in this formula. The factor $\frac{1}{2}$ takes into account the fact that only one half of the conductive electrons, with a suitable spin direction, are strongly scattered by an impurity atom. The factor $(p+1)/p$ takes into account the scatterings with exchange [40].

(×) It is actually somewhat smaller than 5, suggesting that one of the half d shells overlaps slightly the Fermi level, as pictured in Fig. 22. This would explain why the resistivity minimum for Mn is still fairly large.

Further considerations on this topic, in connection with other electrical properties, may be found in reference [29].

2·4. « *Anomalous* » *ferromagnetic alloys in* Ni *and* Co. – Alloys of Ni or Co with solute elements on the left half of the transitional series (Mn, Cr, V, Ti) exhibit strong departures from the simple law

$$\bar{\mu} = \mu_{\text{matrix}} - Zc\mu_{\text{B}} , \tag{18}$$

valid for elements nearer to the matrix elements in the periodic table, as explained in Part I.

These elements have nuclear charges much smaller than the matrix (large negative Z's). They introduce a perturbation V_p which is strongly repulsive as well as highly localized. Now one expects such a perturbation, when strong enough, to subtract a *bound state* from the full half of the d band and to move it towards higher energies. When it comes above the Fermi level E_M, such a state will empty itself into the conduction band, mainly in the other half of the d band with opposite spin direction, because of its high density $n(E_M)$ (Fig. 23). The degeneracy of the d state makes it possible to have room for at most 5 electrons in it. 5 electrons less with spin upwards, turning into 5 more electrons with spin downwards, make a change of $10\mu_{\text{B}}$ when the bound state crosses the Fermi level. Thus the new equation for the average magnetic moment would be [35]:

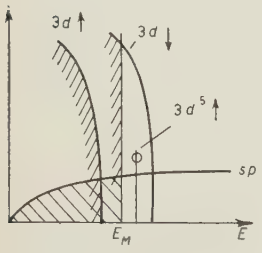


Fig. 23. – Emptying of a 3d bound state in alloys such as **Ni**Cr.

$$\bar{\mu} = \mu_{\text{matrix}} - (Z+10)c\mu_{\text{B}} . \tag{19}$$

Such a law is approximately followed by these anomalous alloys. Thus writing $\bar{\mu} = \mu_{\text{matrix}} - zc\mu_{\text{B}}$, the following table shows that the experimental values of $z - Z$ for Ni alloys go suddenly from values near to 0 up to Mn, to values near to 10 after that element:

Alloy		Z	$z - Z$
Ni	Co	– 1	– 0.2
	Fe	– 2	– 0.8
	Mn	– 3	+ 0.6
	Cr	– 4	8.4
	V	– 5	10.2
	Ti	– 6	9.8

For what concerns the *spatial* distribution of the magnetic moments, the simplest assumption would be to think that the impurity, with its empty half d shell, has a large magnetic moment antiparallel to that of the matrix; the matrix would be left unaffected, as in the normal ferromagnetic alloys like **Ni**Fe (cf. Part I). This model is ruled out by neutron diffraction on **Co**Cr, which shows Cr to have a small, slightly positive magnetic moment, and to reduce strongly that of Co [20] (cf. Fig. 24). This surely comes from the fact that the bound state is fairly *extended*: when appearing above the d band, it starts with an infinite radius, and has probably still a large radius when crossing the Fermi level E_M, because the distance of E_M to the top of the full d band is fairly small, of the order of the magnetic energy kT_c. We picture then the impurity center as in Fig. 25:

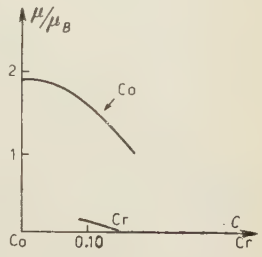


Fig. 24. - Most probable variation of atomic magnetic moments of Co and Cr in **Co**Cr alloys.

a) The full d band with spins upwards has an empty bound state above the Fermi level, which has locally subtracted 5 electronic charges from the band. These missing charges have roughly the same radial extension as the bound state (they are positive holes captured in an empty acceptor level).

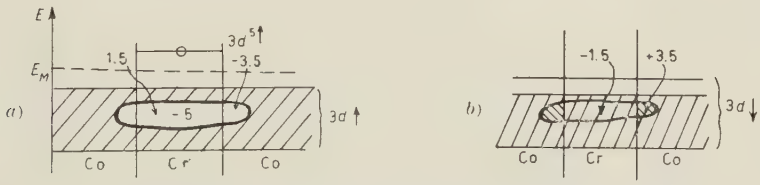


Fig. 25. – Electronic densities in the two halves of the d band for **Co**Cr. (*a*, spins upwards; *b*, spins downwards).

b) The other half of the d band, with spin downwards, has a large density $n(E_M)$, thus screens out easily all the unbalances of charges: on the neighbouring atoms of the matrix, it compensates for the electronic charge missing in the other spin direction; on the impurity atom, it also compensates for the excess nuclear charge Z.

In **Co**Cr, the neutron diffraction results are in agreement with the charge distribution pictured in Fig. 25, where a little more than $\frac{2}{3}$ of the charge of the bound state is on the neighbouring Co atoms.

Finally, the table above shows that, in Ni, the d bound state seems to become empty very suddenly when one goes from Mn to Cr. This shows that

the state is fairly *sharply defined in energy*. Various causes for broadening may be considered:

1) Coulomb correlations.

2) Lattice symmetry.

3) Resonance broadening.

The two first factors are probably not very effective because the state is rather extended (*). As for resonance broadening, contribution from the other half of the d band with opposite spin only arises through the weak spin orbit coupling; resonance with the sp conduction band is fairly weak, because the Fermi energy of this band is small, as it contains only 0.5 electrons per atom.

(*) As it has the full degeneracy $l=5$, at least in Ni alloys, it is nevertheless far enough from the band edge (or localized enough) to mix up contributions from the various component d bands.

REFERENCES

[1] N. F. Mott and H. Jones: *Metals and Alloys* (Oxford, 1936).

[2] L. F. Bates and S. J. Leach: *Proc. Phys. Soc.*, **69** B, 997 (1956); J. A. Catterall: *Phil. Mag.*, **2**, 491 (1957).

[3] E. Vogt and R. Lingelbach: private communication.

[4] C. J. Smithells: *Metals Reference Book* (London, 1949).

[5] The reasoning put forward here is only valid for the shearing of a Brillouin zone without change of volume. The situation is more complicated for a dilatation. Cf. H. Jones: *Proc. Roy. Soc.*, A **147**, 400 (1934); J. B. Goodenough: *Phys. Rev.*, **89**, 282 (1953).

[6] W. Hume Rothery and O. V. Raynor: *Str. of Metals and Alloys* (London, 1954).

[7] A. B. Pippard: to be published.

[8] J. A. Rayne: *Phys. Rev.*, **108**, 22 (1957); P. L. Smith: to be published.

[9] J. O. Linde: *Dissertation Stockholm* (1939).

[10] T. R. Long and C. S. Smith: *Acta Met.*, **5**, 200 (1957); E. I. Salkowitz, A. I. Schindler and E. W. Kammer: *Phys. Rev.*, **105**, 887 (1957) and ibid., to be published.

[11] B. G. Childs and J. Penfold: *Phil. Mag.*, **2**, 389 (1957).

[12] N. F. Mott and R. W. Gurney: *Ionic Solids* (Oxford, 1948).

[13] J. Friedel: *Phil. Mag.*, **43**, 153 (1952).

[14] W. D. Knight: *Adv. Sol. State Physics*, **2**, 93 (1956).

[15] J. Friedel: *Journ. Phys. et Rad.*, **16**, 444 (1955).

[16] O. Kubaschewski and E. L. Evans: *Metallurgical Thermochemistry* (London, 1951).
[17] N. H. March: to be published.
[18] D. Lazarus: *Phys. Rev.*, **93**, 973 (1954).
[19] J. Friedel: *Adv. Phys.*, **3**, 446 (1954).
[20] C. G. Shull: *Congrès Solvay, Bruxelles* (1954).
[21] P. Gombás: *Stat. Theor. des Atoms* (Vienna 1949); J. C. Slater: *Phys. Rev.*, **81**, 385 (1951).
[22] S. Shinohara: *Prog. Theor. Phys.*, **14**, 40 (1954).
[23] M. Laue (von): *Ann. der Phys.*, **44**, 1197 (1914).
[24] H. Fröhlich and F. R. N. Nabarro: *Proc. Roy. Soc.*, A **175**, 382 (1940); M. A. Ruderman and C. Kittel: *Phys. Rev.*, **96**, 99 (1954); N. Blömbergen and T. J. Rowland: *Phys. Rev.*, **97**, 1679 (1955).
[25] N. Blömbergen: *Can. Journ. Phys.*, **34**, 1299 (1956).
[26] N. F. Mott and H. S. W. Massey: *Atomic Collisions* (Oxford, 1949).
[27] H. Brooks: private communication.
[28] C. Zener: *Phys. Rev.*, **81**, 440 (1951).
[29] J. Friedel: *Can. Journ. Phys.*, **34**, 1190 (1956).
[30] L. Roth: *Thesis* (1956).
[31] K. Huang: *Proc. Phys. Soc.*, **60**, 161 (1948).
[32] L. I. Schiff: *Quantum Mechanics* (New York, 1955).
[33] H. Jones: *Phil. Mag.*, **44**, 907 (1953).
[34] E. W. Pugh *et al.*: *Phys. Rev.*, **105**, 814 (1957).
[35] G. F. Koster: *Stuttgart Colloq.* (1956). The idea was first put forward in a qualitative way by the author ([19], p. 492).
[36] V. Heine: *Proc. Phys. Soc.*, A **69**, 505 (1956).
[37] B. Gale and J. Trotter: *Phil. Mag.*, **2**, 1 (1956). Previous references may be found in this paper.
[38] R. A. Ferrell: *Rev. Mod. Phys.*, **28**, 308 (1956); E. Daniel: *Journ. Phys. Chem. Solids*, **4**, 111 (1958); under press (1958).
[39] L. E. Drain: to be published (1958).
[40] J. Friedel: *Journ. Phys. et Rad.* (1958) under press.

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

The Transition from the Metallic to the Non-Metallic State.

N. F. MOTT

Cavendish Laboratory - Cambridge

1. – Introduction.

In all text books on the theory of metals and indeed in this course of lectures you will find detailed discussions of the movement of electrons in a periodic field. These discussions, based on Schrödinger's equation, have been applied with great success to the behaviour of electrons in semiconductors. Here the interaction between the different electrons in the conduction band is frequently negligible, so that once a suitable potential is chosen one has a « one-body problem », that of the possible energy values of a single particle. When the same method is applied to metals, however, the interaction between the electrons in the conduction band is always strong, and one can only treat one of them as moving in a periodic field by averaging in some way the effect of all the others. It is my aim in this lecture to show you that, successful as this approach has been, there are cases in which it gives wrong results. Though correct in general for metals, there are cases in which it is not possible to treat all the electrons in a crystalline solid as moving in a periodic field.

Let me first describe the conventional theory of solids. One starts with Schrödinger's equation

$$\nabla^2\psi + \frac{2m}{\hbar^2}\{W - V(\boldsymbol{r})\}\psi = 0 \ ,$$

where $V(\boldsymbol{r})$ is a *periodic* potential energy function. The solutions ψ_k are of the Bloch form

$$\psi_k = \exp\left[i(\boldsymbol{k}\boldsymbol{r})\right] u_k(\boldsymbol{r}) \ , \tag{1}$$

where u_k is periodic with the period of the lattice. The energy W_k plotted

against $\boldsymbol{k}$ has the form shown in Fig. 5*b*, and one introduces the concept of the density of states $N(W)$. The values of $\boldsymbol{k}$ and allowed energy values $W_{\boldsymbol{k}}$ are divided into « Brillouin zones», of which the allowed energies may or may not overlap.

A. H. WILSON [1] was the first to point out that, if a Brillouin zone is fully occupied, the electrons occyuping the states which make up this zone can make no contribution to an electric current. He was thus able to explain the sharp distinction which exists in nature between pure crystalline materials which at low temperatures show good (metallic) conductivity and those which are insulators. Insulators are substances in which all zones are either full or empty; metals are those in which one or more zones are partly full.

Twenty-six years after WILSON's paper, it seems that he was very nearly right. It is true that the electrons in metals can be described by wave functions of type (1), with values of $\boldsymbol{k}$ *partly* filling one or more Brillouin zones. It is also true that if one describes electrons in a solid by wave functions of type (1) and it then turns out that all zones are full or empty, the substance will be found to be an insulator at low temperatures; typical insulators such as sodium chloride or diamond are to be described in this way. The point that I want to make in these lectures, however, is that there is a class of substances which, if described according to Wilson's scheme, should be metallic conductors and are in fact insulators. Most of the antiferromagnetic oxides come into this class. And not only are there substances of which this can be said, but groups of electrons within substances, for instance the $4f$ electrons within a rare earth metal.

These lectures, then, will develop the basic theory of such systems of electrons and apply them to three types of problems, which are

1) Impurity-band conduction in valence semiconductors;
2) The electrical properties of the antiferromagnetic oxides;
3) The nature of the magnetic electrons in the rare earth and transition metals.

The lectures are based on a series of papers by the present author [2-6], who welcomes the opportunity to make a collected story of them.

2. – Conducting and non-conducting states.

In order to consider the conditions under which the Bloch-Wilson theory breaks down, let us consider an array of hydrogen or other monovalent atoms, and let us denote by a the interatomic distance. Let us discuss in some detail how one would describe the electrons in such an array, and how our description

depends on the interatomic distance. An intuitive approach is the following. If a is large, so that the overlap between neighbouring atoms is small, one should use what is called in the literature the London-Heitler-Heisenberg approach. That is to say, each electron is described by an *atomic* wave function $\psi_n(\boldsymbol{r})$, where

$$\psi_n(\boldsymbol{r}) = \psi(\boldsymbol{r} - \boldsymbol{R}_n) ,$$

ψ being the wave function of a free atom and $\boldsymbol{R}_n$ the co-ordinate of the n-th atom. To describe the wave function of the whole system one sets up a product

$$\psi_1(\boldsymbol{r}_1)\psi_2(\boldsymbol{r}_2) \ldots ; \tag{2}$$

after multiplying by appropriate spin functions one then sets up a Slater determinant antisymmetrical in the co-ordinates of the electrons. This Slater determinant we shall write

$$\Psi_{nc}(q_1, q_2, \ldots) . \tag{3}$$

The q_n are the electronic co-ordinates including spin; *nc* stands for non-conducting. The system described by such a wave function is obviously non-conducting; electrons can change places (electron exchange), but no net current is possible. The question before us is, as a is decreased, how are « polar states » introduced into the wave function in such a way as to allow a current. We shall return to this question below.

First, however, let us look at the metallic description which is appropriate for small a. Here all the electrons are « free », and one may set up a Slater determinant

$$\Psi_c(q_1, q_2, \ldots) \tag{4}$$

from the functions ψ_k corresponding to occupied states. One knows that the use of this Slater determinant suggests that the chance of a given electron being on a given atom is independent of whether another electron of opposite spin is there already (*). This cannot be true, and a great deal of effort from Wigner to Bohm and Pines has been spent on a proper description of « correlation », i.e. the tendency of electrons with opposite spins to keep away from each other. In the free electron model, each electron is seen as moving with a correlation hole round it, in which another electron is less likely to be.

As the distance a increases, this picture does not seem so satisfactory.

(*) Electrons of the same spin are kept off the same atom by the wave function being antisymmetrical.

The wave functions ψ_k can now be given by the tight-binding approximation

$$\psi_k = \sum_n \exp\left[i\boldsymbol{k}\boldsymbol{R}_n\right]\psi_n(\boldsymbol{r})\,. \tag{5}$$

If overlap is small, no energy will be gained by allowing two electrons to sit on one atom; the correlation hole—as Professor BROOKS put it in his lectures—no longer moves with the electron; one needs instead a correlation term which simply prevents an atom from having two electrons on it. Under these conditions it seems very forced to use the function Ψ_c at all; Ψ_{nc}, which takes account of correlation in the zero order of approximation, represents a much more natural approach.

Our aim then is to investigate the transition from wave functions of type Ψ_{nc} to type Ψ_c as a is decreased. Though of course Ψ_{nc} will gradually deviate from the simple determinant described as the overlap increases, we shall give several arguments to show that *the change from a non-conducting to a conducting state* must be sharp; either *none* of the electrons are free to move, or *all* are.

To show this, it must first be emphasized that a wave function of type Ψ_{nc} is non-conducting *only if the number of electrons per atom is unity* (or an integer when the ground state has orbital degeneracy). If we take an electron out, the « hole » left behind can move about and carry a current; the only effect of increasing a is to increase its effective mass, not to stop it moving (*). Similarly if an extra electron is added, so that one atom carries two electrons, this extra electron can move from atom to atom.

Now as *two* hydrogen atoms move together, it is known from the theory of the hydrogen molecule that more and more of these polar terms appear in the wave function. The same must be true for the crystalline array. But the presence of polar states will not necessarily make the system conducting. A new feature turns up in the theory of the infinite lattice which is absent in the theory of the molecule; in the former we must distinguish between *free* carriers, and positive and negative carriers bound together as they are in an exciton. The distinction does not occur for the molecule.

The present author has emphasized in the papers cited [2-5] that at the absolute zero of temperature it is not possible to have a *small* number of free carriers; one must either have none at all or a large number, as in a metal. The argument is the following: a pair of carriers attract each other with a Coulomb force derived from the potential $-e^2/\varkappa r$. They can therefore combine together to form a bound state, similar to an exciton, in which neither

(*) In an ionic lattice the formation of a «polaron» may do this if the effective mass is high. In our discussion, however, we consider the lattice points as fixed.

can carry a current. Therefore a small number of carriers is not possible at $T = 0$.

If, however, the number n of carriers is large—of the order of the number of electrons in a metal—two free carriers attract each other with a force derived from a *screened* Coulomb potential $(-e^2/r)\exp[-qr]$. The quantity q increases with N, and if q is large enough the potential hole becomes too small to allow a bound state to form and the carriers will not combine. This is the main argument that there should be a discontinuous transition from none to a finite density of free carriers.

3. – Wannier functions.

We have explained in the last section that, for a current to be possible, the wave function must include terms that represent free carriers, and also that the existence of free carriers means that polar states are included in the wave function; none the less, we have stated that the existence of polar states does not *necessarily* imply any conductivity. This may be illustrated by describing the non-conducting wave function Ψ_{nc} in terms of Wannier functions. Wannier functions are a set of orthogonal functions formed from the functions ψ_k of any one Brillouin zone, by the transformation

$$w_n(\boldsymbol{r}) = \sum_k \exp[i\boldsymbol{k}\boldsymbol{R}_n]\psi_k(\boldsymbol{r}) ,$$

the summation being over *all* the states in a zone. The function $w_n(\boldsymbol{r})$ is localized in the neighbourhood of the atom $\boldsymbol{R}_n$; if the ψ_k are « tight-binding » functions of type (5), $w_n(\boldsymbol{r})$ reduces to the atomic function $\psi_n(\boldsymbol{r})$. This is only a good approximation when the overlap is negligible, and in other cases a Wannier function is as illustrated in Fig. 1, overlapping onto surrounding atoms.

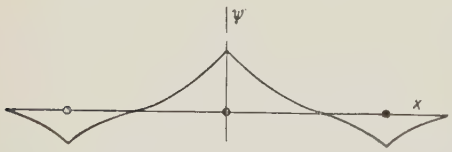


Fig. 1. - Schematic representation of a localised « Wannier » wave function ψ. The sites of atoms are marked with dots.

Suppose now that the lattice is antiferromagnetic so that a definite spin direction can be associated with each atomic site. We can then set up a Slater deteminant Ψ_{nc} from the Wannier functions multiplied by the appropriate spin functions. We note

a) That Ψ_{nc} is formed from *all* the ψ_k of a zone, in contrast to Ψ_c which is formed of half the functions of a zone. The two functions Ψ_{nc} and Ψ_c are thus quite different in form. (On the other hand, if we have *two* electrons per

atom. the determinants formed from ψ_k and w_n are of course identical, and both represent insulating states.)

b) Ψ_{nc} is a real function and apart from states involving spin waves, which do not carry a current, there is no other state of the whole system within an energy of atomic order (a few electron volts) from it. Ψ_{nc} thus describes an *insulating* state. If, however, there is overlap between the atoms, there are polar states (the possibility of two electrons in one atom) as Fig. 1 shows. One can thus have polar states without conductivity.

4. – Slater's model of a non-conducting state.

SLATER [7] has made a suggestion as to how a non-conducting state can be described by wave functions of the Bloch type (1). The assumption is made that the lattice is antiferromagnetic. Any electron described by a Bloch wave function will, in terms of the Hartree-Fock equation, be repelled considerably less by electrons with spin in the parallel direction than by electrons with spins in the antiparallel direction. Therefore in a crystal in the antiferromagnetic state each electron will move in a crystalline field on which is superimposed that of a superlattice. This field of a superlattice will split the first zone of a cubic lattice into two, each containing $\frac{1}{2}N$ states for each spin direction, where N is the total number of atoms. With one electron per atom, the first zone could be completely full and the second empty. The material would then be an insulator.

While this model certainly gives a possible first approximation for an antiferromagnetic substance, it is misleading in one respect. It suggests incorrectly that the material will show metallic conductivity above the Néel point. or in other words that the activation energies required to destroy the long-range antiferromagnetic order and to produce current carriers are of the same order of magnitude, both depending on the width of the energy gap. In fact the former will depend on some overlap integral giving the strength of the exchange or super-exchange interaction, which will tend to zero as a tends to infinity. The activation energy for conduction, on the other hand, tends to $I - E$, where I is the ionization potential of each atom and E the electron affinity.

5. – The interatomic distance at which the transition occurs.

No satisfactory theory of this has been given. The present author [4] has made a rough estimate based on Slater's model, with the assumption that the lattice is antiferromagnetic. For this purpose let us consider a body-

centred cubic lattice and denote the side of the unit cell by b. The atoms on one simple cubic lattice will all have the same spin direction; we shall describe them by the Bloch tight-binding wave functions (5) with energies

$$W = -2\gamma(\cos k_x b + \cos k_y b + \cos k_z b) ,$$

where (k_x, k_y, k_z) is the wave vector, γ an overlap integral. This band is full for electrons of the spin considered; we have to ask whether the top of the band, with energy 6γ, overlaps the bottom of the band obtained when an *extra* electron is brought onto the other half lattice occupied by electrons of opposite spin. If the depth of the bottom of this band is $-U$, the condition for conductivity is

$$6\gamma + U \geqslant I - E ,$$

where I is the ionization energy for one of the atoms, and E the electron affinity. We might equate U to γ', where γ' is an overlap integral for the extra electron moving from one otherwise neutral atom to another; this overlap will, however, be large and we prefer to set for the wave function of the extra electron a plane wave, so that $-U$ is the averaged potential energy of the electron.

Thus

$$\gamma = \int \psi^*(\boldsymbol{r})\psi(\boldsymbol{r} - \boldsymbol{R})(e^2/r)\,\mathrm{d}\tau = (me^4/2\hbar^2)(1 + b/a_{\mathrm{H}}) \exp[-b/a_{\mathrm{H}}] ,$$

for the hydrogenic form of ψ (BETHE [8]), where a_{H} is the hydrogen radius; also

$$-U = \int V(r)\,\mathrm{d}\tau \Big/ \int \mathrm{d}\tau ,$$

where $V(r)$ is the potential within a hydrogen atom so that

$$V(r) = -e^2\left(\frac{1}{a_{\mathrm{H}}} + \frac{1}{r}\right)\exp\left[-\frac{2r}{a_{\mathrm{H}}}\right],$$

and the integration is over a sphere of volume $\frac{1}{2}b^3$. Neglecting small terms containing the exponential this gives

$$U \cong 4\pi e^2 a_{\mathrm{H}}^2/b^3 .$$

If we take $E \sim 0.1I$, as for hydrogen, this gives as the condition for con-

ductivity

$$\frac{4\pi e^2 a_{\mathrm{H}}^2}{b^3} + \frac{bme^4}{2\hbar^2}\left(1 + \frac{b}{a_{\mathrm{H}}}\right)\exp\left[-\frac{b}{a_{\mathrm{H}}}\right] \geqslant 0.9\,\frac{me^4}{2\hbar^2}\,.$$

A numerical solution gives

$$b/a_{\mathrm{H}} \cong 4.5\,.$$

A convenient way of expressing this is to introduce the ratio X of the radius of the atomic sphere round each atom to a_{H}, so that the number c of atoms per unit volume is

$$c = 2/b^3 = 3/4\pi X^3 a_{\mathrm{H}}^3\,.$$

The critical value of X at which the transition to the metallic state should occur is then given by

$$X \cong 2.2\,.$$

Thus we expect the transition to occur when the mean distance between atoms is about four times the mean radius of any one of them.

This derivation is unsatisfactory for many reasons. It neglects the basic principle which we have advanced in favour of a *sharp* transition—namely that a *small* number of holes and electrons cannot exist at $T = 0$. This would suggest that the overlap at the transition point may be somewhat greater than estimated. A better approach might be simply to calculate energies with a metallic wave function, using half the ψ_k of a band and allowing for correlation, and from an antiferromagnetic arrangement of localised orbitals, and to see at what value of a the energies cross. I doubt if it will be possible to do this until we have a method of estimating the effect on energy of correlation in the case of *narrow* bands (large a).

The evidence put forward in the next section does in any case suggest that, as a is decreased and a transition to metallic conductivity occurs, *all* the electrons at once become free.

6. – Impurity-band conduction in semiconductors.

By this we mean conduction by an impurity-activated semiconductor without exciting the electrons into the conduction band. The subject is of very great interest for the purpose of these lectures, because one can vary the concentration of impurity centres in a semiconductor, and thus study the electrical properties of a (random) array of one-electron centres (with properties

similar to one-electron atoms) as a function of the mean distance between them.

Consider for instance an n-type semiconductor in which the concentration of impurity centres is small, so that their wave functions do not overlap. Let I denote the energy necessary to remove an electron from an impurity centre to the conduction band. Then at temperature T the number of electrons in the conduction band is proportional to $\exp[-\frac{1}{2}I/kT]$, so that the resistivity will tend to infinity as T tends to zero.

As is well-known the « radius » of the wave functions of these hydrogen-like orbits is very large, being of order $\varkappa\hbar^2/m_{\text{eff}}e^2$, and considerable overlap may be expected at concentrations of about 10^{17} cm^{-3}. In this region, then, we may expect a transition from a conducting or metallic state for the electrons in the impurities to a non-conducting state in which temperature is essentially necessary for conduction. Fig. 2 shows some recent observations by FRITSCHE [9] of the conductivity at 2.5 °K of p-type germanium activated by gallium. It will be seen that for distances between the centres greater than about 200 Å the conductivity is only about one thousand times less than for a metal; the resistivity is in fact almost independent of temperature in this range. It is clear that the impurity centres are here close enough together for their electrons to form a conducting or metallic assembly, in other words a condensed Fermi gas. The mean free path depending little on T will be a consequence of the disordered arrangement of the centres—a much more random arrangement than in a liquid metal.

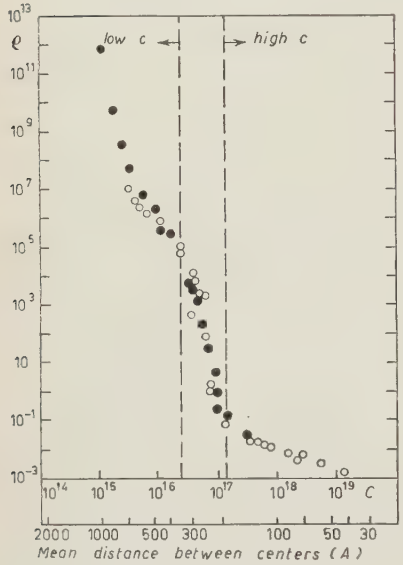


Fig. 2. – Conductivity ϱ in Ωcm of p-type germanium at 2.5 °K as a function of concentration c in cm^{-3} of the majority carriers.

The drop in the resistivity as c approaches 10^{17} cm^{-3} is very rapid. Our theory suggests that the activation energy for conductivity should drop *discontinuously* from a finite value to zero as c is increased, at any rate for all crystalline arrangements of atoms; the absence of a discontinuous drop in Fig. 2 could well be explained by the random arrangement of the centres. Unfortunately there is no way of testing this hypothesis, since a crystalline arrangement cannot be obtained.

The analysis given in Sect. **5** shows that the activation energy should drop as c is increased; the type of curve to be expected is thus of the type shown

in Fig. 3. It is a marked feature of the materials in the « high c » region of Fig. 2 that the conductivity and Hall constant are very constant from say 20 to 2.5 °K, suggesting that *all* the carriers are free; it does seem a question of all or none.

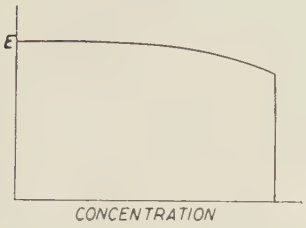


Fig. 3. – Activation energy E for conduction as a function of concentration, according to the theory put forward in these lectures.

The nature of the conductivity on the low c side in Fig. 2 is of great interest. The type of conductivity observed [10] for three concentrations in the region 20 to 1.5 °K is shown in Fig. 4. For high c, as we have said, the resistivity is independent of T. For lower values of c, in the region 20 to 5 °K we have the normal temperature-activated conductivity with electrons moving in the conduction band and the activation energy dropping slightly as c approaches the critical value. In the low temperature region below about 5 °K the slope goes over to a lower value. It appears that the electrons are *not* excited into the conduction band but move (by tunnel effect) from one impurity centre to another.

In the author's opinion (compare reference [10] for some experimental evidence) this can *only* occur through the presence of a certain number of *minority* centres. Suppose that the majority centres be donors and the minority acceptors. Then the acceptors will *all* acquire electrons from donors, so that some of the donors will be full and others empty. Under these conditions it is perfectly legitimate to speak of an impurity band, *partly* occupied, below the conduction band. The band rapidly becomes narrower and the effective mass greater as the distance between the impurity centres increases.

However, even in this case an activation energy for movement is expected. Each acceptor carries a negative charge. The carriers may be thought of as « holes » in a nearly full donor impurity band, and each one will be trapped by a charged acceptor. The activation energy to free it should be of order [9]

$$e^2/(\tfrac{1}{2}\varkappa r_0)\,, \tag{6}$$

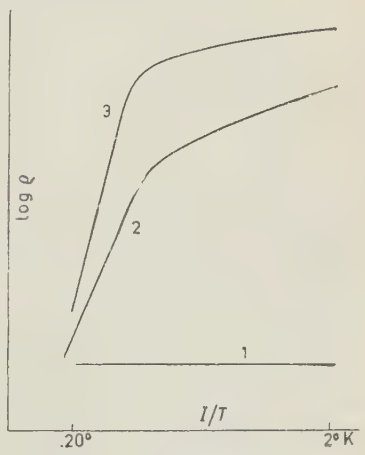


Fig. 4. – Schematic behaviour of resistivity ϱ of impurity-activated valence semiconductors. 1) High concentration (conducting or metallic wave-functions). 2) Medium concentration; some overlap but no conduction without thermal activation or compensation. 3) Low concentration with slight overlap.

where r_0 is the mean distance between the donor centres. We think that the slope in Fig. 4 between 5 and 2 °K may correspond to this activation energy.

The recent paper by KOENIG and GUNTHER-MOHR [10] discusses in detail the experimental evidence relevant to the above hypothesis about impurity band conduction. In particular, formula (6) seems to give a fair figure for the relevant activation energy.

7. – Electrical properties of oxides.

It was the fact that the antiferromagnetic oxides and similar compounds do not show metallic conductivity, either above or below the Néel point, which first suggested the necessity of introducing non-conducting states into the theory. Consider for example the case of nickel oxide (NiO). If this is thought of as built up of nickel ions Ni^{++} and oxide ions, it is clear that, if the electrons are described by Bloch wave functions, the $3d$ band of nickel will be partly occupied. This conclusion remains valid even if, as is certainly the case, the $2p$ oxygen orbitals overlap the nickel ion considerably, making use of the empty $3d$ orbitals to do so. The number of states in the $3d$ band of nickel is not affected by the behaviour of the wave functions of the $2p$ band of oxygen in the neighbourhood of the nickel atoms.

The absence of conductivity in NiO is certainly not due to the distance between the metal ions being too great to allow conductivity by tunnel effect; this is shown by the fact that NiO activated by Li_2O, for example, and thus containing Ni^{+++} ions, is a semiconductor. The « hole » represented by Ni^{+++} among the other Ni^{++} ions *is* able to move from atom to atom.

It seems reasonable to suppose, therefore, that the electrons in the inner shells of the nickel ions, to which the antiferromagnetism is due, are to be described by « non-conducting » wave functions of the type described, and this is why they do not show metallic conductivity, either above or below the Néel point. It is interesting to contrast the much higher conductivity of Fe_3O_4, an oxide containing a *mixture* of ions of two types, Fe^{++} and Fe^{+++}.

HEIKES [11] has given an interesting table, reproduced here, showing the electrical and magnetic properties of oxides, sulphides etc., of the transition metals; this is shown in Table I, together with the ratio of the radius r of the metal ion (Pauling's estimate) to the inter-ionic distance R. It will be noticed that, while compounds with five or more electrons in the $3d$ shell are antiferromagnetic and are non-metals, those with four are metallic conductors and show ferromagnetism. There is not much difference in the values of r/R for the two classes and it seems most unlikely that the conductivity is due to any large overlap between $3d$ electrons leading to a metallic type of $3d$ band. It is much more probable that the overlap between the full shells of

Te^{--} ions, in CrTe for example, is great enough to lead to metallic conduction there.

TABLE I. – *Magnetism and electrical properties of binary compounds (due to* HEIKES).

	Substance	Resistivity (ohm) (room temp.)	Temperature coefficient of resistivity	$\frac{r}{R}$
Ferromagnetic (4 electrons in *d* shell)	MnBi	10^{-4}	positive	0.207
	MnAs	10^{-3}	positive	0.217
	MnSb	10^{-3}	positive	0.214
	CrTe	10^{-4}	positive	0.273
Antiferromagnetic (5 or more electrons in *d* shell)	MnO	10^{9}	negative	0.25
	FeO	10^{8}	negative	0.243
	CoO	10^{8}	negative	0.291
	NiO	10^{8}	negative	0.238
	CuO	10^{10}	negative	—
	Fe_2O_3	10^{10}	negative	—
	MnS	10^{5}	negative	0.218
	MnSe	10^{5}	negative	0.208
	MnTe	10^{2}	negative	0.238

The present writer [4] has made the tentative suggestion that overlap between the anions will be favoured by the presence of less than five electrons in the $3d$-shell of the transition metal for the following reason. If there are less than five electrons there will be empty states for both spin directions; therefore the electrons of the outer shell of the ions can enter these low-lying states whatever their spin direction, and thus will spread out further in space.

8. – Incomplete inner shells of transition and rare earth metals.

In considering the incomplete inner shells of metals, it is obvious that similar considerations may be valid; it may be proper to describe these electrons by wave functions of the Bloch type, in which case they will contribute to the conductivity, or by wave functions of the non-conducting type, in which case they will not. But since the conductivity will come in any case mainly from electrons in the outer shell (conduction electrons), observations of other kinds must be used to determine which type of wave function is appropriate. A discussion of these observations will be given in this section.

We must, however, first remark that the argument which we have given for a *discontinuous* change from a state with no free carriers to one with a

finite number does not apply in this case. We have argued that positive and negative carriers attract one another with a force derived from the potential $-e^2/\varkappa r$ and that with such a potential they could always combine to form uncharged pairs. In a metal, however, this potential will always be changed into a screened potential $(-e^2/r)\exp[-qr]$, the screening being provided by the conduction electrons. There is no *a priori* argument against the presence in any inner band of a very small number of free carriers. In fact, one knows from many investigations that bands containing a very small number of carriers do exist in many metals (cf. HEINE [12]). But none the less it seems to us legitimate, in metals too, to ask whether the wave function of the electrons in an inner shell is of conducting or non-conducting type.

In the first place we must emphasize that the difference between the two types only arises if the number of electrons in each shell is *integral*. Fig. 5

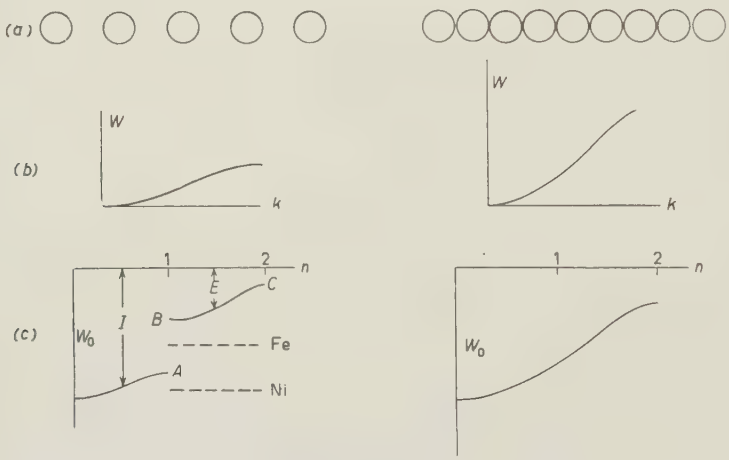


Fig. 5. – *a*) Shows an array of atoms for large and small values of the interatomic distance respectively. *b*) Energy of a single particle as a function of $\boldsymbol{k}$. *c*) Energy W_0 per electron of an assembly of n electrons per atom. If these curves relate to the inner shells of the transition metals nickel and iron, the Fermi surface is as shown by the dotted lines. The assembly is non-conducting in the states A, B, C.

shows this; the diagrams *a*) show the atoms of an array for large and small interatomic distance, *b*) shows the energy W as a function of k for a single particle moving in the field of these atoms, and *c*) the energy W_0 per electron of the whole system as a function of the number n of electrons per atom in the band. It is characteristic of the non-conducting wave function that there will be a gap in the curve *c*) for each integral value of n, which does not occur in the conducting case. When we are considering incomplete inner shells of a metal, we have to ask whether the Fermi surface—i.e. the energy per

electron of a conduction electron there—lies in this gap or not. If it does, then the inner band should contain an integral number of electrons and should make no contribution to the conductivity; if on the other hand the Fermi surface does not lie in the gap, the number of electrons in the inner band will be non-integral and they can contribute to a current.

Turning now to actual metals, it seems certain that for the transition metals cobalt, nickel and palladium the second of these two descriptions is correct. The two former are ferromagnetic and the saturated magnetic moment in terms of Bohr magnetons is non-integral (0.6 and 1.6); since the early work of SLATER [13] and of the present author [14] this has been explained in terms of a partially full *d*-band and a conduction (4*s*) band which overlap. The model has been successful in explaining a number of properties, for instance

a) The magnetic properties of the alloys of these metals with noble and other metals. If nickel, for instance, is alloyed with copper, the extra electron provided by each copper atom goes mainly into the 3*d* band, thus reducing the magnetic moment per nickel atom.

b) The electrical properties of these metals and their alloys. Nickel for instance has considerably higher electrical resistance than copper, silver than palladium. To explain this and other peculiarities of the electrical behaviour, a model was proposed some time ago by the present author [15]. According to this model, most of the current is carried by electrons in the 4*s* (conduction) band, a comparatively small part being carried by the holes in the 3*d* band because of the high effective mass there. However, the scattering of the conduction electrons (the reciprocal of the mean free path) is determined mainly by transitions from the *s* to the *d* band, caused as are other scattering processes by lattice vibrations or impurities. This model is particularly convenient for explaining the electrical resistance of nickel (Fig. 6). It will be seen that the resistance increases rapidly near the Curie temperature. This is because all the holes in the *d* band have the same spin directions, and therefore only half the conduction electrons can suffer these *s*-*d* collisions; the other half will suffer transitions from one state in the *s*-band to another and therefore have a much longer mean free path as in copper. At higher temperatures there will be holes

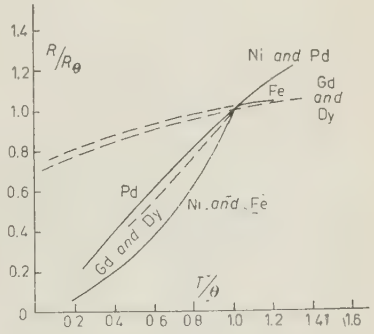


Fig. 6. – Electrical resistance of various ferromagnetic metals: Ni and Pd, GERRITSEN [23]; Fe, POWELL [24]; Gd and Dy, LEGVOLD [25]. Note the similar behaviour above the Curie point of Ni and Pd and of Fe and the rare earths.

with both spin directions in the d-band and therefore electrons with bond spin directions will suffer collisions.

c) X-ray absorption edge. Many transition metals show a strong maximum of the L_3 absorption at the absorption edge; a curve observed by Cauchois [16] for nickel is shown in Fig. 7(*a*). This is

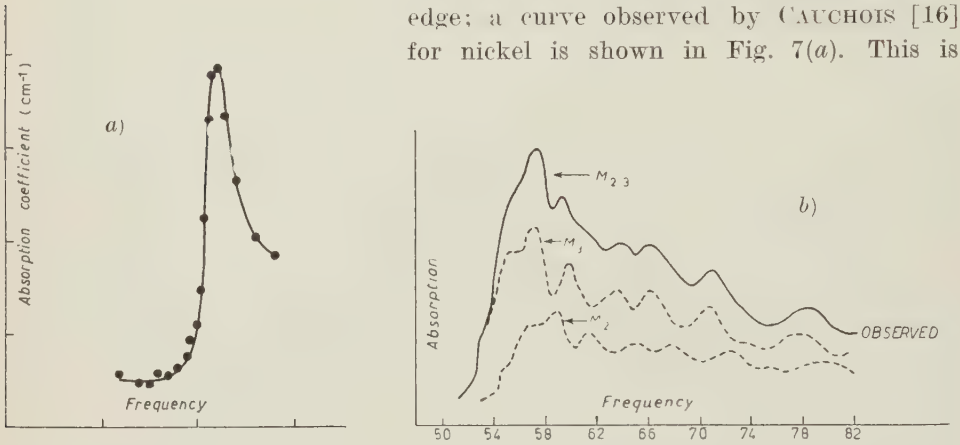


Fig. 7. – X-ray absorption edges for transition metals: *a*) The L_3 edge for nickel observed by Cauchois; the divisions on the frequency scale are 7 eV apart. *b*) The M_2 or M_3 edges of iron, as deduced by Tomboulian from the observations of Carter and Givens (frequency in eV).

supposed due to the vacant states in a d-band in which the density of states is high.

If we turn from nickel and cobalt to the rare earth metals such as gadolinium, it seems unlikely that the $4f$ shells will contain a non-integral number of electrons. The comparatively low Curie temperature suggests a narrow band width, and although in one or more of these metals the Fermi surface might accidentally lie in the middle of such a band, this is hardly likely to be the case for many of them. If this is so, the resistance anomaly of this metal near the Curie point (Fig. 6) must be due to quite a different process, namely the scattering of the conduction electrons by the disordered spins, a process which does not depend on the lattice vibrations at all.

A metal in which there is as yet some uncertainty about the proper description of the magnetic electrons is body-centred iron. Here the magnetic moment per atom is 2.2 Bohr magnetons per atom, and it is a plausible assumption that the inner shell contains just two electrons per atom contributing to the magnetic moment, and that the remainder comes either from orbital moment or from some small magnetization of the conduction electrons (cf. Zener [17]). Griffith [18] suggest that these two electrons are in e_g orbitals, that is in $3d$ orbitals with symmetry $x^2 - y^2$ etc., of which there are

two per atom with each spin direction and which should interact less strongly with their surroundings in a cubic lattice than those with symmetry xy etc.

The present author has suggested that these electrons ought to be described above or below the Curie temperature by a non-conducting wave function and that they contain exactly two electrons per atom.

Consequences of such a hypothesis are:

a) That on alloying with non-transition metals, the magnetic moment per iron atom would not change as it does in nickel, but would remain constant. That this is the case has been emphasized by STONER [19] and by COLES and HUME-ROTHERY [20], and was in fact the most important reason for introducing this model of non-conducting states.

b) The magnetic anomaly in the electrical resistance of iron, which looks very much like that in nickel, must be due to quite a different mechanism, namely the scattering of conduction electrons by disordered spins in the inner $3d$ shells. Certain experiments aimed at distinguishing the two mechanisms are in progress at the present time (for a discussion see MOTT and STEVENS [6]).

c) In any X-ray absorption process in which electrons in an inner p-shell make transitions to the empty states above the Fermi surface the absorption coefficient should have a sharp maximum, due to the empty e_g states, to the short-wave side of the absorption edge. TOMBOULIAN has recently analysed the M absorption spectrum of iron obtained by CARTER and GIVENS [21], in which the initial state is $3p$. This state is a doublet; the result of the analysis by TOMBOULIAN [22] is to give the curve shown in Fig. 7*b* for the absorption due to either state of the doublet. It will be seen that the white line is considerably displaced from the edge.

The evidence for non-conducting states in iron is then fairly strong. A more detailed discussion will be published by the present author elsewhere [6].

9. – Conclusion.

I hope from these examples I may have convinced that the usual band theory of solids based on Bloch wave functions of type (1), though enormously successful, is not of universal validity. It is always legitimate to use it if one has a non-integral number of electrons per atom, but if the number is integral it may predict conductivity and related phenomena in cases when no conductivity is observed. Examples may be drawn from the inner shells of metal atoms both in the metallic state and in oxides and also from impurity-band conduction in valence semiconductors.

REFERENCES

[1] A. H. Wilson: *Proc. Roy Soc.*, A **133,** 450 (1931).
[2] N. F. Mott: *Proc. Phys. Soc. Lond.*, **62** 416 (1949).
[3] N. F. Mott: *Adv. in Metal Physics*, **3**, 76 (1952).
[4] N. F. Mott: *Report of Rugby Meeting on Semiconductors*, London, Physical Society (1956).
[5] N. F. Mott: *Can. Journ. Phys.*, **39**, 1356 (1956).
[6] N. F. Mott and K. W. H. Stevens: *Phil. Mag.* (8) **2**, 1364 (1957).
[7] J. C. Slater: *Phys. Rev.*, **82** 568 (1951).
[8] H. A. Bethe: *Handbuch der Physik*, **24**/1, 529 (1933).
[9] H. Fritzsche: *Purdue University Second Quarterly Report* (1956).
[10] S. H. Koenig and G. R. Gunther-Mohr: *Journ. Phys. Chem. Solids*, **2**, 268 (1957).
[11] R. R. Heikes: *Phys. Rev.*, **99**, 1232 (1955).
[12] V. Heine: *Proc. Roy. Soc.*, A **240**, 340 (1957).
[13] J. C. Slater: *Phys. Rev.*, **41**, 57 (1936).
[14] N. F. Mott: *Proc. Phys. Soc. Lond.*, **47**, 571 (1935).
[15] N. F. Mott.: *Proc. Phys. Soc. Lond.*, **47**, 571 (1935).
[16] Y. Cauchois and C. Bonnelle, *Comptes Rendus*, **245,** 1230 (1957).
[17] C. Zener: *Phys. Rev.*, **81**, 440 (1951).
[18] J. S. Griffith: *Journ. Inorg. Nucl. Chem.*, **3**, 15 (1956).
[19] E. C. Stoner: *Rep. Progress in Physics*, **11**, 43 (1946).
[20] W. Hume-Rothery and B. R. Coles: *Adv. in Physics*, **3**, 149 (1954).
[21] D. E. Carter and M. P. Givens: *Phys. Rev.*, **101**, 1469 (1956).
[22] D. H. Tomboulian: *Technical Report ORD*, 669 (1957).
[23] A. N. Gerritsen: *Handbuch der Physik*, **19**, 187 (1956).
[24] R. W. Powell: *Proc. Phys. Soc. Lond.*, **51**, 416 (1939).
[25] S. Legvold: *Rev. Mod. Phys.*, **25**, 129 (1953).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Electrons and Plasmons.

D. PINES

Princeton University - New Jersey

1. – Introduction.

In these lectures I wish to consider the behaviour of electrons in solids when one takes into account their mutual interaction. I shall pay particular attention to the following questions:

1) What is the ground state energy for the free electron gas, i.e. a system of interacting electrons in a uniform background of positive charge?

2) What are the elementary excitations of electrons in solids?

In discussing the first question, I shall follow the approach which BOHM and I [1] developed for the calculation of the correlation energy for the free electron gas. The correlation energy is defined as the difference in energy between the Hartree-Fock value and any better calculation. It is so called because the Hartree-Fock method takes into account only the « accidental » correlations due to the Pauli principle and neglects the correlations in particle positions brought about by their mutual interaction; any improved calculation will take the latter correlations into account.

In discussing the second question I shall try to summarize the results of some investigations which Mr. NOZIÈRES and I have carried out during the past two years [2]. We shall see that the plasmons, that is, the quantized high frequency plasma oscillations of the electron gas as a whole, are for most solids a well defined elementary excitation. We shall also try to understand to what extent we are justified in regarding « electrons » in solids as independent elementary excitations.

2. – Qualitative considerations.

The Hamiltonian for the free electron gas may be written as (*):

$$H = \sum_i \frac{P_i^2}{2m} + 2\pi e^2 \sum_{\substack{k \neq 0 \\ i \neq j}} (\exp \cdot [i\boldsymbol{k} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j)] / k^2 , \tag{1}$$

where the exclusion of the term with $k = 0$ takes into account the contribution of the uniform background of positive charge. The density fluctuations of the electron gas, ϱ_k, provide a useful tool for obtaining qualitative insight into the behaviour of the system. We have

$$\varrho_k = \int \varrho(x) \exp [- i\boldsymbol{k} \cdot \boldsymbol{x}] \, \mathrm{d}x = \sum_i \exp [- i\boldsymbol{k} \cdot \boldsymbol{x}] , \tag{2}$$

and $\varrho_0 = n$, the electron density. The equation of motion of ϱ_k is:

$$\ddot{\varrho}_k = - \sum_i \left(\frac{k \cdot P_i}{m} + \frac{\hbar k^2}{2m} \right)^2 \exp [- i\boldsymbol{k} \cdot \boldsymbol{x}_i] - \sum_q \frac{4\pi e^2}{mq^2} \, \boldsymbol{k} \cdot \boldsymbol{q} \, \varrho_{k-q} \varrho_q . \tag{3}$$

The first term on the right-hand side represents the influence of the kinetic energy of the particles; the second arises from their mutual interaction. We may rewrite (3) separating out the term with $\boldsymbol{q} = \boldsymbol{k}$ in the second term on the right-hand side, we then have

$$\ddot{\varrho}_k + \omega_P^2 \varrho_k = - \sum_i \left(\frac{k \cdot P_i}{m} + \frac{\hbar k^2}{2m} \right)^2 \exp [- i\boldsymbol{k} \cdot \boldsymbol{x}_i] - \sum_{q \neq k} \frac{4\pi e^2}{mq^2} \, \boldsymbol{q} \cdot \boldsymbol{k} \, \varrho_{k-q} \varrho_q , \tag{4}$$

where

$$\omega_P = (4\pi n e^2 / m)^{\frac{1}{2}} \tag{5}$$

is the plasma frequency.

The ϱ_k oscillate at a frequency ω_P provided we can neglect the two terms on the right-hand side of (4). Let us see under what circumstances we may do so. We may approximate the first term by $k^2 V_0^2 \varrho_k$, where V_0 is the velocity of an electron at the top of the Fermi distribution. Let us furthermore assume that the second term is negligible, an approximation which is equivalent to linearizing the equations of motion of the density fluctuations. It is frequently referred to as the random-phase approximation. The reason our study of the equations of motion of ϱ_k is qualitative is that it is not possible within this framework to justify the random phase approximation.

(*) We are Fourier analysing in a box of unit volume, subject to periodic boundary conditions.

The condition for collective oscillatory behaviour of the ϱ_k, and hence of the entire electron gas, is then, in addition, that

$$\frac{k^2 V_0^2}{\omega_P^2} \ll 1 \qquad \text{or} \qquad k^2 \ll k_c^2 , \tag{6}$$

where k_c^2 is of the order of ω_P^2/V_0^2. Thus in general we may expect that the electron system behaves collectively for $k \ll k_c$. For $k \gg k_c$ it behaves like a system of free individual particles, since then the first term on the right hand side of (3) dominates. In the region of k near k_c the behaviour will be more complicated.

Let us put in some numbers. For a typical electron gas with a density of $\sim 10^{23}$ electrons/cm^3 we find that $\hbar\omega_P \cong 12$ eV, and k_c^{-1} is of the order of the inter-electron spacing. Thus under ordinary circumstances, we do not expect to observe the collective excitation at a frequency ω_P (the plasmons) because the energy to excite a plasmon, $\hbar\omega_P$, is greater than E_0, the energy of an electron at the top of the Fermi distribution (and far greater than kT). The wave length $\lambda_c = k_c^{-1}$ represents the minimum wave length for which we might expect collective behaviour; it also is the effective range of the actual electron interaction. To see this, let us write our basic Hamiltonian (1) as:

$$H = \sum_i \frac{P_i^2}{2m} + \sum_{k \neq 0} \frac{2\pi e^2}{k^2} (\varrho_k^* \varrho_k - n) . \tag{7}$$

Since the ϱ_k for $k < k_c$ exhibit oscillatory behaviour near ω_P we might expect that in some suitable co-ordinate system, they would be described by a set of equivalent harmonic oscillators, and would contribute only a zero-point energy to the system. We are therefore left with an effective Hamiltonian,

$$H^1 = \sum_i \frac{P_i^2}{2m} + \sum_{\substack{k > k_c \\ i \neq j}} \frac{2\pi e^2}{k^2} \exp [i\boldsymbol{k} \cdot (\boldsymbol{x} - \boldsymbol{x}_j)] , \tag{8}$$

a Hamiltonian appropriate to a set of electrons interacting with a screened Coulomb interaction of range k_c^{-1}. In other words, the long range part of the interaction is frozen out because of the high frequency of the plasma oscillations it brings about.

We may reach these conclusions more physically in the following way. As an electron moves, it tends to push other electrons out of its way, because of their mutual Coulomb repulsion. In fact, the electrons are quite efficient in staying out of one another's way. For any given electron, one is not likely to find another electron within a distance of the average inter-electron spacing; each electron is essentially surrounded by a correlation hole of missing electrons.

Outside this hole, one sees then the field of the electron plus the field of the uniform position charge surrounding the electron up to a radius of the order of r_0, where $4\pi r_0^3/3 = 1/n$ is the volume per electron. The net field seen by any other electron at distances greater than r_0 is nearly zero; hence the effective interaction between the electrons is screened, with a screening radius comparable to the inter-electron spacing.

The plasma oscillations describe the correlations in the electron motion as this screening action takes place. If, for instance, there is a charge unbalance at some point in the electron gas the electrons will tend to move in such a way as to screen it out. In so doing, they will move too far, overshooting the equilibrium point somewhat. They are attracted back in, overshoot again, etc., and one has an oscillation at a frequency near ω_P.

Before going into a more detailed consideration of the free electron gas, let us survey the regimes of different electron density. We may characterize the electron density by r_s, the inter-electron spacing measured in units of the Bohr radius a_0. We have

$$\frac{4\pi r_s^3 a_0^3}{3} = \frac{1}{n} . \tag{8}$$

For actual metals r_s varies between about 2 and 5.5, if we regard the valence electrons as making up our free electron gas. There are for the general problem three regions of interest:

1) High densities $r_s \ll 1$

2) Metallic densities $2 \leqslant r_s \lesssim 5.5$

3) Low densities $r_s \gg 10$.

In the high density case, the electronic kinetic energy dominates the scene, and the interaction between the electrons is relatively weak. The collective plasmon modes, n' in number, represent an insignificant fraction of the total number of degrees of freedom. Let us define $\beta = k_c/k_0$, the maximum plasmon wave vector measured in units of the maximum electronic wave vector. The relative number of collective degrees of freedom is

$$\frac{n'}{3n} = \frac{k_c^3}{6k_0^3} = \frac{\beta^3}{6} .$$

As we have seen $k_c \sim \omega_P/V_0$, so that $\beta \sim r_s^{\frac{1}{2}}$. Hence $n'/3n \sim r_s^{\frac{3}{2}}$, and goes to zero with small r_s. Because of the relative weakness of the interaction, in this domain the Coulomb interaction can be handled rigorously, and this is

what GELL-MANN and BRUECKNER [3] have done recently. They find that the energy per electron can be written as a power series in r_s, which has the form

$$E = \frac{a}{r_s^2} + \frac{b}{r_s} + c + d \ln r_s + e r_s + f r_s \ln r_s + \ldots \text{ Ry} , \tag{9}$$

where the first two terms represent the average Fermi energy and the exchange energy, and the remaining terms are the correlation energy. I might add that the basic approximation which they make is the random-phase approximation, which is rigorously satisfied in the very high density limit.

The very low density regime presents a quite different picture. Here, as WIGNER [4] first pointed out, the electrons will be found in a regular array, about which they carry out oscillations. In other words, one not only has n longitudinal plasmon degrees of freedom (as might have been anticipated from our expression for $n'/3n$), but one also has $2n$ transverse oscillatory modes. The energy per-electron may be written as

$$E = \frac{a'}{r_s} + \frac{b'}{r_s^{\frac{3}{2}}} + \frac{c'}{r_s^2} + \ldots \text{ Ry} , \tag{10}$$

where the first term represents the exchange energy plus the leading term in the correlation energy, the second term represents the zero-point energy of the electrons oscillating about their equilibrium positions, and the third term arises from the van der Waals interaction between the electrons. We might call the low density case the « strong coupling » regime, since the interaction between electrons is all important and actually forces them into a regular array. Of course the random phase approximation is no longer valid here, since the electrons are now in a periodic array.

The case of actual metallic densities represents an « intermediate coupling» regime, in which both collective and individual particle behaviour are important. The relative number of collective degrees of freedom arising from the Coulomb interaction is not inappreciable; as we shall see, it is given approximately by

$$\frac{n'}{3n} = 0.01 r_s^{\frac{3}{2}} ,$$

and so varies from 3% to 13% for metals. The random phase approximation turns out to be quite satisfactory for the long wave length oscillatory plasmon modes, but is questionable for the large wave vectors appearing in the remaining screened Coulomb interaction. As a result, it is difficult to obtain a truly accurate value for the ground state energy of the system. We shall

see that the energy can be written as a sum of the following terms:

$$E = \frac{a}{r_s^2} + \frac{b}{r_s} + c'' + d \ln r_s + e'' r_s + \ldots \text{ Ry}, \tag{11}$$

in which the coefficients c'' and e'' may differ somewhat from their counterparts in (9). The accuracy of (11) is estimated to be about 20%.

3. – Correlation energy in the free electron gas.

As we have seen the electrons in a dense electron gas display both collective and individual particle behaviour. Our basic problem is to develop a formalism which allows explicitly for the introduction of n' collective co-ordinates to describe the plasma oscillations, and which still maintains the possibility of describing the motion of individual electrons. Bohm and I treated this problem by introducing n' extra co-ordinates through the addition of certain terms to the basic Hamiltonian, (1). We then related the new co-ordinates to the actual oscillating variables (which for long wave lengths will be close to the ϱ_k) by a series of canonical transformations. While doing this, one can guarantee that the expected Hamiltonian be identical with the original Hamiltonian by imposing a set of n' subsidiary conditions on the extended system wave-function. (This simultaneously guarantees that the total number of degrees of freedom of the extended system be just $3n$). However, as we shall see, the real usefulness of the method stems from the fact that for many purposes the subsidiary conditions can be ignored. We shall show that they do not affect the ground state energy, the plasmons, or individual electron excitations; in other words, they act only to inhibit the simultaneous excitation of many electrons [5].

Let us consider the following family of extended Hamiltonians:

$$H = \sum_i \frac{P_i^2}{2m} + \sum_k \frac{2\pi e^2}{k^2} (\varrho_k^* \varrho_k - n) + \sum_{k<k_c} \left(\frac{\pi_k^* \pi_k}{2} + \pi_k^* \sqrt{\frac{4\pi e^2}{k^2}} \varrho_k \right), \tag{12}$$

with the associated set of subsidiary conditions:

$$\pi_k \Psi = 0 \qquad (k < k_c). \tag{13}$$

The π_k represent a set of field momenta, n' in number, where $n' = k_c^3/6\pi^2$. They commute with the particle co-ordinates and momenta. It is straightforward to show that if the ground state of our extended Hamiltonian (2) is non-degenerate, the ground state wave function will automatically satisfy the subsidiary conditions, (13). This follows because the subsidiary conditions are

not invariant under a translation, while the Hamiltonian (12) is invariant (*). Such will be the situation for any extended Hamiltonians which have a lower bound, as is obviously the case with (12).

We have, therefore, a new kind of variational approach. We are able to vary the form of the Hamiltonian with which we work, as well as the wave-functions appropriate to a given Hamiltonian. We cannot in so doing obtain a lower energy than the true ground state energy of the free electron gas, since the entire family of extended Hamiltonians (12) possess that ground state energy in common. In the remainder of this section, then, we shall consider what is the best extended Hamiltonian to use (in other words, the choice of k_c) and what is the best ground state energy we thereby obtain. We do not need to concern ourselves with the subsidiary conditions until we take up the excitations of the electron system.

Thus far all of our considerations have been purely formal; the π_k are quite without physical significance. We begin to come to grips with the physics of the problem when we carry out a canonical transformation which relates the π_k to the actual oscillating variables. We first carry out the following transformation:

$$(14)\qquad \begin{cases} \pi_k \to \pi_k - \sqrt{\dfrac{4\pi e^2}{k^2}}\,\varrho_k\,, \\ Q_k \to Q_k\,, \\ P_i \to P_i - i\sum\limits_{k<k_c} \sqrt{\dfrac{4\pi e^2}{k^2}}\,\boldsymbol{k}_i \exp\left[-i\boldsymbol{k}\cdot x_i\right] Q_k\,, \\ x_i \to x_i\,, \end{cases}$$

Q_k is the collective co-ordinate which is canonically conjugate to π_k. We obtain the following family of Hamiltonians:

$$(15)\qquad H = \sum_i \frac{P_i^2}{2m} + H_{\text{s.r.}} + \sum_{k<k_c} \left\{ \frac{\pi_k^*\pi_k}{2} + \omega_P^2 \frac{Q_k^* Q_k}{2} - \frac{2\pi n e^2}{k^2} \right\} + H_{\text{int}} + U\,.$$

(*) The proof goes as follows. Define the translation operator $U(\Delta X)$, such that $U^{-1}X_i U = X_i + \Delta X$; $U^{-1}\varrho_k U = \varrho_k \exp\left[-i\boldsymbol{k}\cdot\Delta X\right]$. We may also require that $U^{-1}\pi_k U = \pi_k \exp\left[i\boldsymbol{k}\cdot\Delta\boldsymbol{X}\right]$ and $U^{-1}Q_k U = Q_k \exp\left[-i\boldsymbol{k}\cdot\boldsymbol{X}\right]$. We then have $U^{-1}H_{\text{ext}}U = H_{\text{ext}}$ where H_{ext} is any of the family of Hamiltonians, (12). Let E be the non-degenerate lowest state of H_{ext} and Φ its corresponding eigenstate. Since π_k and $U(\Delta X)$ both commute with H_{ext}, Φ is also a simultaneous eigenstate of these operators, so that

$$\begin{aligned} H_{\text{ext}}\Phi &= E\Phi\,, \\ \pi_k\Phi &= \beta_k\Phi\,, \\ U(\Delta X)\Phi &= \alpha\Phi\,. \end{aligned}$$

We further have $U(\Delta X)\pi_k\Phi = \alpha\beta_k\Phi$ by combining the last two equations. We also have

$$U(\Delta X)\pi_k\Phi = \left(U(\Delta X)\pi_k U^{-1}\right)U\Phi = \alpha\beta_k \exp\left[i\boldsymbol{k}\cdot\Delta\boldsymbol{X}\right]\Phi\,,$$

so that $\beta_k = \beta_k \exp\left[i\boldsymbol{k}\cdot\Delta\boldsymbol{X}\right]$. Since this is true for arbitrary ΔX we must have $\beta_k \equiv 0$. The theorem applies, in fact, to any non-degenerate eigenstate of H_{ext}.

$H_{s.r.}$ represents a short-range interaction between the electrons, of range k_c^{-1};

$$H_{s.r.} = \sum_{\substack{k>k_c \\ i\neq j}} \frac{2\pi e^2}{k^2} \exp\left[i\boldsymbol{k}\cdot(\boldsymbol{x}_i - \boldsymbol{x}_j)\right], \tag{16}$$

H_{int} describes a linear interaction between the electrons and plasmons,

$$H_{int} = i \sum_{k<k_c}{}^{i} \left(\frac{k\cdot P_i}{m} + \frac{\hbar k^2}{2m}\right) \sqrt{\frac{4\pi e^2}{k^2}}\, Q_k \exp\left[-i\boldsymbol{k}\cdot\boldsymbol{x}_i\right] \tag{17}$$

and U describes an electron-plasmon interaction which is bi-linear in the plasmon co-ordinates,

$$U = \frac{2\pi e^2}{m} \sum_{\substack{k,q\,i \\ k\neq q}} Q_k^* Q_q \sqrt{\frac{1}{k^2 q^2}}\, \boldsymbol{q}\cdot\boldsymbol{k} \exp\left[i(k-q)\cdot x_i\right]. \tag{18}$$

If we neglect H_{int} and U we see that we have in (15) a system of electrons interacting via a screened Coulomb interaction of range k_c^{-1}, and a set of n' plasmons, all at frequency ω_P. This is the picture we were led to by our consideration of the equations of motion of the density fluctuations. In fact the neglect of H_{int} corresponds to neglecting the first term on the right-hand side of (4), while the neglect of U is just the random phase approximation. Now, however, we are in position to estimate how valid these approximations are.

We may do this simply provided we are justified in regarding $H_{s.r.}$ as a relatively small perturbation on the motion of the electrons; as we shall see this requires that k_c be sufficiently large. We may then neglect $H_{s.r.}$ when considering the effect of H_{int} and U on the electron-plasmon system. It turns out that the coupling introduced between the electrons and plasmons by H_{int} may be measured by the coupling constant:

$$g^2 = \left\langle \left(\frac{k\cdot P_i}{m\omega_P}\right) \right\rangle_{av} \cong \frac{\beta^2}{2r_s}, \tag{19}$$

where the average is over all electron momenta and values of k less than k_c. (One determines this by writing down the perturbation-theoretic canonical transformation which eliminates H_{int} to first order. g^2 represents the expansion parameter which appears in that transformation (see BP - III [1] and the following section)). For sufficiently small values of β, then, H_{int} represents a relatively weak coupling between the electrons and the plasmons, which can readily be treated by perturbation-theoretic methods. U turns out to represent an even weaker electron-plasmon coupling.

We may calculate the size of U by second order perturbation theory: we

find a change in energy per electron which is

$$\Delta E = -\sum_n \frac{|U_{n0}|^2}{E_n - E_0} \cong \frac{\beta^7}{48}\hbar\omega_P .$$

The ratio of ΔE to the average kinetic energy furnishes a measure of the importance of U; it is

$$\frac{\beta^4}{96}\frac{\hbar\omega_P^5}{3E_0} \cong \frac{\beta^7}{30} r_s^{\frac{1}{2}} ,$$

and is clearly negligible for all metallic densities.

Let us proceed to calculate the ground state energy of the electron gas. As a first approximation, let us assume the electron-plasmon and electron-electron interaction are small. The wave function of the system is then

$$\Phi = \Psi_{\text{osc}}\chi_0 , \tag{20}$$

where Ψ_{osc} is a product of simple harmonic oscillator wave functions, one for each collective mode, describing the plasmon ground state and χ_0 is the usual Slater determinantal wave function appropriate to the ground state of a gas of free electrons. The energy of the system is

$$E = \frac{3}{5}nE_0 + \sum_{k<k_c}\left(\frac{\hbar\omega_P}{2} - \frac{2\pi ne^2}{k^2}\right) + \langle H_{\text{s.r.}}\rangle_{\text{av}} , \tag{21}$$

where the first term is the kinetic energy, the second the difference between the zero-point oscillations of the plasmons and the self-energy of the charge distribution they have replaced, and the last is the exchange energy associated with $H_{\text{s.r.}}$. One finds for the energy per electron,

$$E = \frac{2.21}{r_s^2} - \frac{0.916}{r_s} - 0.458\frac{\beta^2}{r_s} + 0.019\frac{\beta^4}{r_s} + 0.866\frac{\beta^3}{r_s^{\frac{3}{2}}}\ \text{Ry}. \tag{22}$$

The last three terms in (22) represent the correlation energy in this approximation, since the first two are the system energy in the Hartree-Fock approximation. This part of the correlation energy arises from the fact that we have replaced the long-range part of the Coulomb interaction by the plasmons; it is just

$$\sum_{k<k_c}\left\{\frac{\hbar\omega_P}{2} - \frac{2\pi e^2}{k^2}\langle\varrho_k^*\varrho_k\rangle_{\text{av}}\right\} .$$

One method for choosing β consists in minimizing the system energy (22).

If we keep only the leading terms in β^2 and β^3, we get a quite simple expression for β,

$$\beta = 0.353 r_s^{\frac{1}{2}} .$$

A somewhat better value for β is obtained if we wait to minimize the energy until after we include the contribution from H_{int}; the latter may by calculated using second-order perturbation theory, and is, per electron

$$(\delta E)_{\text{int}} = 0.708 \frac{\beta^5}{r_s^{\frac{5}{2}}} \left(1 + \frac{3}{10}\beta^2\right) - 0.517 \frac{\beta^4}{r_s^2} - 0.058 \frac{\beta^6}{r_s^2} \text{ Ry} . \tag{23}$$

The new value for β, obtained by minimizing the sum of (22) and (24) (again keeping only the leading terms), is

$$\beta \cong 0.4 r_s^{\frac{1}{2}} . \tag{24}$$

The correlation energy, with this choice of β, is then

$$E_{\text{corr.}}^{\text{l.r.}} = -0.024 + 0.0006 r_s \text{ Ry} . \tag{25}$$

With the choice of β in (24) we are in quite good shape with respect to the consistency of our assumptions about the smallness of H_{int} and U. The coupling constant, (19), is

$$g^2 = 0.08 ,$$

so that the corrections beyond the second-order from H_{int} are certainly negligible. The contribution to the correlation energy from U is very small, being

$$(\delta E)_U \cong -0.00003 r_s^2 \text{ Ry} . \tag{26}$$

The remaining contributions to the correlation energy come from $H_{\text{s.r.}}$. These contributions have thus far only been calculated using conventional second-order perturbation theory for the many-electron problem. The dominant contribution comes from the interaction between electrons of antiparallel spin. One finds (see SSP [1]) for the correlation energy with these interactions

$$E_{\text{corr.}}^{\text{s.r.}} = -\{0.0254 - 0.0622 \ln \beta + 0.00637 \beta^2\} \text{ Ry} . \tag{27}$$

There is also a contribution from the interaction between electrons of parallel spin. The integrals which appear in the latter expression are very much more

complicated, and it has only recently been possible to obtain a value for it from some Monte Carlo calculations carried out for GELL-MANN and BRUECKNER [4]. One finds:

$$E_{\text{corr.}}^{\text{exch.}} = +0.021 + 0.0622 \ln \beta - 0.021\beta^2 \text{ Ry} . \tag{28}$$

If we now substitute the value for β from (24), we find

$$E_{\text{corr.}}^{\text{s.r.}} = -\{0.035 - 0.0311 \ln r_s + 0.0010 r_s\} \text{ Ry} \tag{29a}$$

$$E_{\text{corr.}}^{\text{exch.}} = -\{0.081 - 0.0311 \ln r_s + 0.0034 r_s\} \text{ Ry} . \tag{29b}$$

The total correlation energy is the sum of (25) and (29); it is

$$E_{\text{corr.}} = -0.140 + 0.0622 \ln r_s - 0.0038 r_s \text{ Ry} . \tag{30}$$

It is of interest to compare our calculation of the correlation energy with the earlier calculation of WIGNER [6] and the recent calculation of GELL-MANN and BRUECKNER [4]. To compare with WIGNER we should really consider the value for the correlation energy obtained by neglecting the second-order correlations between electrons of parallel spin, that is, neglecting $E_{\text{corr.}}^{\text{exch.}}$. That value is

$$E_{\text{corr.}}^{1} = -0.105 + 0.0311 \ln r_s - 0.0004 r_s \text{ Ry} . \tag{31}$$

It agrees with the calculation of WIGNER,

$$E_{\text{corr.}}^{\text{Wig}} = -\frac{0.88}{r_s + 7.8} \text{ Ry} , \tag{32}$$

to better than 6% over the region of actual metallic densities. Such a close agreement is not accidental, as has been shown recently by KREISMENT [7], who shows that WIGNER's assumed wave function is very close to our wave function (20), when one applies the subsidiary condition to (20).

GELL-MANN and BRUECKNER obtain the following value for correlation energy

$$E_{\text{corr.}}^{\text{G-M,B}} = -0.096 + 0.0626 \ln r_s + 0(r_s) \text{ Ry} . \tag{33}$$

Their result is only valid for extremely high densities ($r_s \ll 1$). Our expression (30), is only in fair agreement with the result (33); I believe the discrepancy may be traced to the inadequacy of second-order perturbation theory in our calculation of the correlation energy arising from $H_{\text{s.r.}}$. From the

comparison of (30) with (33), and from various numerical estimates of the validity of perturbation theory (see SSP [1]), I am led to estimate the accuracy of our expression (30) as about 20% (*).

One final word on the choice of β. One might ask why it makes sense to choose β by minimizing only a part of the correlation energy. The answer, I believe, is that we have chosen the lesser of various evils. We could not include the short-range part of the correlation energy, because that is not known accurately, and would in fact lead us to a divergent result. We have chosen a value of β which makes the most of that portion of the correlation energy we can calculate accurately, that associated with the long wave length momentum transfers. In so doing, we have fair accuracy in treating $H_{s.r.}$:

A somewhat more *ad hoc* approach, which might in fact lead to a more accurate answer, would be to take a rather larger value of β than (24); an increase to $\beta = .6 r_s^{\frac{1}{2}}$, for instance, would not lead to an appreciably different long-range correlation energy, nor less accurate one, but might produce a more accurate value for the short-range correlation energy. One point which should be kept in mind involves the physics of the situation; β gives the maximum wave-vector for which the plasmon may be regarded as an independent elementary excitation, and that requires that $\beta \leqslant 0.47 r_s^{\frac{1}{2}}$, since for this value a plasmon can spontaneously decay into a free electron.

4. – The transformation to independent collective modes.

In our Hamiltonian (15), we have a set of n' plasmons, in relatively weak interaction with the electrons (provided of course, that we choose k_c properly). The weakness of the interaction makes possible a systematic treatment of its effects, and leads to an effective isolation of the collective modes of the system, the plasmons. We can carry out a canonical transformation which, to any desired order in g^2, decouples the plasmons from the electrons. Let us see how this looks, without going into any details:

(*) The fact that our « intermediate » coupling result, which is reasonably accurate for actual metallic densities, has the same structure as the high density expansion (and with only a small correction of order r_s) might lead one to assume that the Gell-Mann and Brueckner result is valid for actual metallic densities. This is almost certainly not the case. The expression, (30), or the result of WIGNER, (32), when combined with the other terms in a cohesion energy calculation (see the lectures by Prof. BROOKS) yields rather good agreement with experiment. The Gell-Mann and Brueckner expression, (33), would give cohesion energies which were some 13 kcal/mole too low (not enough cohesion); the remaining terms in the calculation of the cohesion energy are not in error by such an amount.

1) Divide the Hamiltonian (15) into three parts:

$$H_0 = \sum_i \frac{P_i^2}{2m} + \sum_{k<k_c} \frac{P_k^* P_k}{2} + \frac{\omega^2 Q_k^* Q_k}{2},$$

$$H' = \sum_{k<k_c} (\omega_P^2 - \omega^2) \frac{Q_k^* Q_k}{2}.$$

We are neglecting U, and H_{int}, which means we neglect the coupling between different momenta in the electron-electron and electron-plasmon interaction; ω is the frequency of the plasmons after we have isolated the plasmons, and is to be determined.

2) Consider the transformation $\Psi = \exp[iS/\hbar\Phi]$; $H \to \exp[-iS/\hbar] \cdot H \exp[+iS/\hbar]$ which expanded, reads:

$$H = H - \frac{i}{\hbar}[S, H] - \frac{1}{2\hbar^2}[S, [S, H]] + \dots .$$

Choose S such that

$$-\frac{i}{\hbar}[S, H_0] = -H_{\text{int}} .$$

The expansion in powers of S is a perturbation-theoretic expansion. H_{int} is proportional to g, the coupling constant for the electron-plasmon interaction, and S is therefore also of order g (H_0 being of order unity, in this sense).

3) In the new Hamiltonian the lowest order corrections to H_0 will be of order g^2; they arise from

$$-\frac{i}{2\hbar}[S, H_{\text{int}}] .$$

If we make the random phase approximation in treating these terms, we find that one such is proportional to $Q_k^* Q_k$, and that we may define ω in such a way as to thereby cancel H'. (The correction to the plasmon frequency from the interaction is thus of order g^2.) The new dispersion relation is

$$1 = \frac{4\pi e^2}{m} \sum_i \frac{1}{(\omega - (k \cdot P_i/m))^2 - \hbar^2 k^4/4m^2},$$

which may be expanded, in powers of g^2, as

$$\omega^2 = \omega_P^2 + \frac{3}{5} k^2 V_0^2 + \frac{\hbar^2 k^4}{4m^2} + \dots .$$

The remaining terms in $-(i/2\hbar)[S, H_{\text{int}}]$ describe a new effective interaction between the electrons, which has come about because the electrons are coupled to the plasmons through H_{int}. It is of order g^2 compared to the original unscreened Coulomb interaction, describes a long-range screened interaction between the electrons, and is given by:

$$H_{\text{r.p.}} = -\frac{\pi e^2}{m^2} \sum_{\substack{k<k_c \\ i,\ j}} \frac{[k\cdot(P_i - \hbar k/2)]\,[k\cdot(P_j + \hbar k/2)]}{k^2\omega\,(\omega - (k\cdot P_j/m) - \hbar^2k^2/2m)} \exp[ik\cdot(x_i - x_j)] + \text{c. c.}$$

4) Finally we remark that the subsidiary conditions now involve only the electron co-ordinates (again to order g^2). The old subsidiary conditions were

$$\left(\pi_k - \sqrt{\frac{4\pi e^2}{k^2}} \sum_i \exp[-ik\cdot x_i]\right)\psi = 0 \qquad (k < k_c).$$

These are transformed to

$$\sum_i \frac{\omega^2}{\omega^2 - ((k\cdot P_i/m) - \hbar k^2/2m)^2} \exp[-ik\cdot x_i]\psi = 0 \qquad (k < k_c). \tag{34}$$

5) There remain terms in the Hamiltonian (and the subsidiary condition) which are of order $g^2\cdot H_{\text{int}}$, and which could be eliminated to give rise to corrections of order g^4. The basic structure of the new Hamiltonian,

$$H = H_0 + H_{\text{rp}} \tag{35}$$

and the new subsidiary conditions (34), are unchanged, and the g^4 corrections are in practice negligible.

5. – Plasmons.

In our Hamiltonian (34), the plasmons are an *independent mode of excitation* of the electron gas. Plasmon excitation will occur when a charged particle of energy greater than $\hbar\omega_p$ passes through the electron gas. It is the dominant mode of energy transfer for wave-vectors less than k_c (*). The

(*) The interaction between the charged particle and the electron gas is $\sum_k (4\pi e^2/k^2)\cdot\varrho_k \exp[-ik\cdot R_0]$ (R_0 is the particle position) which transforms for wave vectors less than k_c to, approximately $\sum_{k<k_c} \sqrt{4\pi e^2/k^2}\,\pi_k \exp[-ik\cdot R_0]$, from which the cross-section for plasmon excitation may be calculated.

plasmons may be damped only through the terms in the Hamiltonian we have thus far neglected, those arising from a coupling between H_{int} and $H_{\text{s.r.}}$ which formally appears via $[S, H_{\text{s.r.}}]$. Damping cannot occur through H_{int} directly, because as long as $\beta \lesssim .47 r_s^{\frac{1}{2}}$ energy and momentum cannot be conserved in the interaction ; i.e.,

$$\hbar\omega > \frac{k \cdot P_i}{m} + \frac{k^2\hbar}{2m} .$$

The situation in an actual solid is a bit different. In the presence of a periodic potential, one finds not only intra-band transitions for the electrons, which are fairly well described by a free electron gas model, but also inter-band transitions, which fall quite outside the scope of a free electron gas approach. One might think that the inter-band transitions (which may, after all, occur for a wide spectrum of energies bracketing the plasmon energies) would markedly alter any conclusions based on a free electron gas model; actually, they do not, as was first pointed out by MOTT [8].

MOTT showed, using a semi-classical argument, that if the majority of the inter-band transitions correspond to excitation frequencies, ω_{no} which are small compared to ω_p, then the plasmons will be little affected by the inter-band transitions. Because the plasmon energies are so high ($\hbar\omega_p \sim 15$ eV for a wide variety of solids) such a situation is frequently encountered in practice.

Recently, NOZIÈRES and I (*) have carried out a treatment of electron interaction in solids which closely parallels that I have described for the free electron gas. I should like to summarize the essential points of the treatment:

1) The first canonical transformation (14), which relates the plasmon variables to the ϱ_k, goes through in exactly the same way in the presence of a periodic potential, since the transformation leaves that potential unchanged. Therefore the basic Hamiltonian is that of (15), with the addition of a periodic potential.

2) All the consequences of that potential for the plasmons and the electron interaction may be described in terms of the one-electron excitation frequencies ω_{no} for electrons moving in that potential, and the one electron oscillator strengths, $f_{\text{no}}(k)$. Consider the Hamiltonian, $H_B = \sum_i (P_i^2/2m) + V(\boldsymbol{r}_i)$ where $V(\boldsymbol{r}_i)$ is the periodic potential. The eigenfunctions Φ_n of H_B are determinants of Bloch waves φ_n. We are interested in matrix elements of ϱ_k between various states of H_B, as for instance $(\varrho_k)_{\text{no}}$; (ϱ_k) induces transitions from the ground state to excited states which differ from the ground state in that a

(*) See references [2], where a list of references to earlier work along these lines may also be found.

single electron has made a transition from a state KV (to use the reduced zone scheme) to another state $K+k$, V', where $V'=V$ corresponds to an intra-band transition, $V'\neq V$ to an inter-band transition. ω_{no} is the excitation frequency $(E_n-E_0)/\hbar$ for that transition. The oscillator strengths for the transitions to excited state are then defined by

$$f_{no}(k)=\frac{2m}{\hbar k^2}\omega_{no}|(\varrho_k)_{no}|^2. \tag{36}$$

They satisfy the sum rule, $\sum_n f_{no}(k)=n$, where n is the number of valence electrons.

3) The condition for the existence of plasmons as an independent mode of excitation is that the electron gas be highly polarizable, which means that a sizeable fraction of the oscillator strengths correspond to $\omega_{no}\ll\omega_P$. The plasmons may be isolated by a suitable canonical transformation when this is true; such isolation is also possible where the majority of the oscillator strenghts correspond to frequencies much higher than ω_P.

4) The plasmon dispersion relation is

$$1=\frac{4\pi e^2}{m}\sum_n\frac{f_{no}}{\omega^2-\omega_{no}^2}, \tag{37}$$

which reduces to $\omega^2\cong\omega_P^2$ when the majority of the oscillator strengths correspond to $\omega_{no}\ll\omega$.

5) A generalization of the relation (37), allows one to include the effect of the core polarizability on valence electron plasmons. In the case that the vast majority of the core electrons have important excitation frequencies large compared to the plasmon frequency ω, we may treat core electrons and valence electrons simultaneously in deriving the plasmon dispersion relation. One finds

$$1=\frac{4\pi e^2}{m}\left\{\sum_n\frac{f_{no}}{\omega^2-\omega_{no}^2}+\sum_n\frac{F_{no}}{\omega^2-\Omega_{no}^2}\right\}, \tag{38}$$

where the F_{no} and Ω_{no} are the core electron oscillator strengths and excitation frequencies. If we now make use of the expression for the core electron dielectric constant at frequency, ω,

$$\varepsilon=1+4\pi\alpha=1+\frac{4\pi e^2}{m}\sum_n\frac{F_{no}}{\Omega_{no}^2-\omega^2},$$

we may re-write (38) as

$$1=\frac{4ke^2}{m\varepsilon(\omega)}\sum_n\frac{f_{no}}{\omega^2-\omega_{no}^2}, \tag{39}$$

which is exactly what one would expect from a semi-classical argument. Further, for the case under consideration, $\varepsilon(\omega) \simeq \varepsilon_0$, the static core dielectric constant.

6) I should like to emphasize that these conclusions are not confined to metals; they apply equally well to valence electrons in a semi-conductor and even to many insulators (those with small energy gaps). What is relevant for an understanding of the importance of Coulomb interactions is a comparison of ω_P, the plasma frequency calculated assuming the electrons in question are free, and $\bar{\omega}_{no}$ the average one-electron excitation frequency. Where $\omega_{no} \gg \omega_P$ as might be the case for molecular crystals (the solid rare gases), the Coulomb interaction is small, may be treated by perturbation theoretic techniques, and no plasmons will be found. On the other hand, for the vast majority of solids $\bar{\omega}_{no} < \omega_P$. For such solids, Coulomb interaction is important and may be described in terms of the plasmons, which constitute a well defined mode of excitation of the system.

7) There are two ways of determining the plasmon energy experimentally. One is by the study of the characteristic energy loss spectrum of electrons passing through thin solid films. The pioneer experiments in this field were by RUTHEMANN [9] and LANG [10]; the invention of electrostatic focusing techniques has greatly stimulated work on these lines. This determination is not always reliable, but it has been carried out for a wide variety of solids [11]. A second method involves the study of optical properties of solids. When the frequency of an electromagnetic wave reaches the plasma frequency, the solid passes from a reflecting to a transmitting region [12]. The frequency involved is in the far UV; as a result experiments have been carried out thus far only for the alkali metals.

I have discussed in some detail elsewhere [13] the experimental evidence for the existence of plasmons in solids, and I here wish only to summarize some of the results on the plasmon energies. My summary takes the form of the following five tables (taken from NP III) with a few remarks appended to each. In all of these tables, $\hbar\omega_P$ is the calculated value for the plasmon energy, assuming all the valence electrons are free, and $\hbar\omega_{el}$ is the experimentally observed value, which is taken from reference [13].

TABLE I.

Element	Be	Mg	Al	B	C	Si	Ge
$\hbar\omega_P$ (eV)	19	11	16	24	25	17	16
$\hbar\omega_{el}$ (eV)	19	10	15	19	22	17	17
ε				6.2	5.7	12	16

Table I: These are « simple » solids, in that the valence electrons are weakly bound, and the core electrons are strongly bound. ε is the dielectric constant obtained from optical data. The good agreement between ω_{el} and ω_P is gratifying. The shift away from the free electron value for B and C may be understood from the fact that the valence electrons in these solids are somewhat more tightly bound, as is reflected in their lower electronic polarizabilities.

Table II.

Element	Li	Na	K	Rb	Cs
$\hbar\omega_P$ (eV)	8.1	6.0	4.4	4.0	3.6
$\hbar\omega$	8.0	5.7	3.9	3.4	2.9
$\hbar\omega_{opt}$	8.02	5.91	3.94	3.65	3.27

Table II: One would likewise expect the alkali metals to be « simple» solids, and they are; $\hbar\omega$ is the plasmon energy when the core polarizability is taken into account, while $\hbar\omega_{opt}$ is the optically determined value. When the effect of core electron polarizability on the plasmon energy is taken into account, the agreement between theory and experiment is even better.

Table III.

Element	Ti	Cr	Mn	Fe	Co	Ni	Cu	Zn
$\hbar\omega_P$ (eV)	17	24	28	31	34	35	36	32
$\hbar\omega_{el}$	22	24	22	21	21	23	20	23

Table III: No treatment at present is satisfactory for the transition metals. If we regard the plasmons as composed of both s and d electrons, we are including many electrons which are tightly bound. For these electrons $\bar{\omega}_{no} \gg \omega_P$ so that we overestimate ω_{el}. On the other hand, the polarizability of the d electrons is enormous, and at these energies one can certainly not consider separately the s electrons. The behaviour of the plasmon energy from Ti through Mn may be understood qualitatively in the following way. For Ti, the valence electrons have an average excitation frequency $\bar{\omega}_{no}$ somewhat below ω_P, so that $\omega_{el} > \omega_P$. For Mn, certainly the reverse is true. At Cr we appear to have a cross-over, in that there are as many transitions with $\omega_{no} > \omega_P$ as $\omega_{no} < \omega_P$; the result is a broad line at nearly ω_P. A similar effect appears to occur for other metals with 6 valence electrons (Mo, W, Se).

TABLE IV.

Element	Ag	Cd	In	Sn	Sb	Te
$\hbar\omega_P$ (eV)	9	11	11	12	24	15
$\hbar\omega_{el}$	23	20	12	12	15	18

Table IV: Here we observe a shift, at In, from metals in which the core electrons are weakly bound, and take part in the plasmons, to those in which they may be regarded as tightly bound.

TABLE V.

Compond	ZnS	PbS	BeO	MgO	Al_2O_3	SiO_2	SnO_2	KBr	KCl	NaCl
$\hbar\omega_P$ (eV)	17	16	29	25	27	25	26	13	14	16
$\hbar\omega_{el}$	17	15	29	25	23	25	20	13	13	16
ε	5.6	15	3.0	3.0	3.1	2.2	4	2.4	2.2	2.4

Table V: The good agreement between the free electron value and the experimentally observed value for such a wide variety of semi-conductors and insulating compunds is a bit surprising, particularly so for those solids which have a rather low electronic polarizability.

6. – Effective electrons in solids.

I should now like to consider the present basic theoretical justification for the individual-particle like elementary excitations in solids («effective» electrons). We all know how very crude the usual one-electron approximation in solids is; we also know how very successful it is. Indeed, experiment demands that there be something very like an independent particle excitation as evidenced, for instance, by the de Haas-von Alphen effect in metals and cyclotron resonance in semi-conductors. The question is rather one of whether, in the light of our present knowledge of the influence of electron interaction in solids, we can understand the effective electrons. In what follows, I shall be summarizing a paper which Mr. NOZIÈRES and I have written on the subject [2]. Our principal conclusion is that good qualitative arguments may now be given for the effective electrons, but that a detailed quantitative justification remains to be made.

Let us first assume that we have no long-range interaction between the electrons. We thus consider the structure of the eigenstate spectrum of $H_0^B + H_{s.r.}$

where $H_{s.r.}$ is of sufficiently short-range that no collective excitations exist. We may then follow and extend an approach recently given by GELL-MAN [14] and by LANDAU [15] for the free electron gas.

1) The eigenstates of H_0^B alone are Slater determinants built with Bloch functions, $\varphi_{k\nu}$. In the ground state Φ_0, all levels inside a certain surface S in k space are filled; those outside are empty. S is the Fermi surface. An excited state Φ_n is described by p electrons outside S ($K_1 \ldots K_p$), and p holes inside $K_1' \ldots K_1'$); its energy is

$$E_n = E_0 + \sum_{i=1}^{p} E(K_i) + \sum_{i=1}^{p} E(K_i') \,. \tag{40}$$

2) Switch on $H_{s.r.}$, allowing the charge to increase progressively from 0 to e. Assume that in so doing the energy levels vary continuously from their free electron values, so that an eigenstate $\Phi_n(e)$ may be followed continuously from $\Phi_n(0)$. There will then be a lowest state $\Phi_0(e)$ which defines a perturbed Fermi surface S'; an excited state $\Phi_n(e)$ is likewise deduced from the unperturbed $\Phi_n(0)$.

3) Consider the energy of a state $\Phi_n(e)$ with one excited electron K_i, and one hole K_i', $E(K_i K_i')$. We may write:

$$E_{K_i K_i'} - E_0 = (E_{K_i'} - E_0) + (E_{K_i K_i'} - E_{K_i'}) \,, \tag{41}$$

where $E_{K_i'}$ is the energy of the system with the hole alone. The first term on the right-hand side is then the energy to create a hole, while the second is the energy to create an electron in the presence of the hole. The latter will depend on the hole only if the interaction between the hole and the electron gives rise to a strong spatial correlation in their positions (e.g., a bound state corresponding to an exciton). Otherwise we may assume the hole and electron are spread throughout the solid, and their mutual interaction will be of order $1/N$ compared to their interaction with the remaining electrons. We may then write:

$$E_{K_i K_i'} - E_0 = E(K_i) + E(K_i') \,, \tag{42}$$

where $E(K_i)$ and $E(K_i')$ are the energies of an « effective » electron and hole which are independent.

4) Where (42) is valid, it may be generalized to an arbitrary number, N, of electrons and holes, provided $\hat{N} \ll N$. The energy spectrum is then identical in form with (40). We may define a Fermi surface for our « effective » electrons, and we may go on to calculate the density of states, specific heat, spin susceptibility, etc. I should add that such calculations have been carried out for the $H_{s.r.}$ given by (16) and yield good agreement with experiment [16].

How far may these conclusions be extended to actual electrons in solids? The long-range part of the Coulomb interaction does introduce strong spatial correlations between the electrons, which correspond to the plasmons. However, as long as we restrict our attention to ordinary temperature ($KT < E_F$) or low energy excitations ($E_{exc} < \hbar\omega_P$), plasmon excitation will not be important, and we might hope that effective electrons do exist. Let us then consider the modifications brought about by the long range part of the Coulomb interaction.

The basic Hamiltonian, after the decoupling of the plasmons from the electrons is

$$H_0^B + H_{s.r.}^{\mathfrak{S}} + H_{pl} + H_{r.p.},$$

where H_{pl} is the plasmon Hamiltonian, and $H_{r.p.}$ describes the new effective long-range electron interaction. We also have a set of subsidiary conditions which act only on the electrons,

$$\Omega_K \Psi = 0 \qquad (K < K_c).$$

The complications brought about by the Coulomb interaction then reside in $H_{r.p.}$ and Ω. I should like to indicate what these are and how we may account for them in at least qualitative fashion.

1) Normalization. Care must be taken in normalizing the extended system wave functions when we take into account the subsidiary conditions as is evident from the original form of the subsidiary condition,

$$\pi_K \psi = 0.$$

2) $H_{r.p.}$ gives rise to just those correlations in the electronic positions which bring about the satisfaction of the subsidiary conditions; such correlations involve many electrons, and an explicit treatment of them has not yet been carried out. Thus if we treat $H_{r.p.}$ by second-order perturbation-theory, we find an additional energy which is comparable to that calculated in lowest order, both being very small. It is therefore necessary to go beyond perturbation-theory to understand $H_{r.p.}$, though the improvement in energy so obtained will not be significant.

3) The subsidiary conditions affect the electronic degrees of freedom. Actually they do so only to a very limited extent. They do not inhibit low energy excitations of the system, but play a role only for excitations comparable to the plasmon energy. What they tell us is that we cannot describe the plasmons twice. Even long wavelength high energy ($> \hbar\omega_p$) individual-

electron excitations may not be affected, since at such wave-lengths the plasmons involve the simultaneous excitation of many electrons.

In other words the subsidiary conditions act only to influence that class of excitations which correspond to the simultaneous excitation of many electrons with a total excitation energy greater than $\hbar\omega_p$. The preceding statement no longer applies for $k \leq k_c$ since then the plasmon states begin to involve only one electron.

Nozières and I have attempted to make more quantitative the foregoing qualitative conclusions regarding the role of the long-range part of the Coulomb interaction. In so doing, we have met with only limited success. If we assume that we deal only with long-range interactions (neglecting $H_{s.r.}$), we may then treat rigorously the interaction of a minority group of electrons (say those which are near the surface and which are susceptible to temperature excitation) with the majority group (the remaining electrons). We may, in fact, decouple the minority electrons from the plasmons, from the subsidiary conditions and from the majority electrons. We thus arrive at a set of independent effective electrons, not subject to subsidiary conditions, which correspond to « dressed » electrons, each surrounded by a cloud of virtual plasmons and majority electron excitations, in such a way that their effective interaction is characterized by the dielectric constant of the solid.

If we now try to take into account $H_{s.r.}$ we find that its principal effect will be to shift the energy levels of these electrons at the top of the Fermi distribution. We do not at present understand how to take these fairly large shifts into account after we have treated the long-range part of the interaction. Also, it does not appear feasible to take $H_{s.r.}$ into account first and then treat the long-range part of the interaction, because the procedure outlined above for treating $H_{s.r.}$ has not been reduced to Hamiltonian form.

To sum up, for a metal we can define effective electrons as long as we have only a long-range interaction, or only a short-range interaction. It is certainly likely that effective electrons exist in the presence of both parts of the Coulomb interaction, but a formal proof that they exist is still lacking.

Finally, I should like to mention that the situation appears rather brighter for conduction electrons in a semi-conductor or an insulator, because in these cases the band-gap may be expected to reduce the energy shifts associated with $H_{s.r.}$. A perturbation theoretic treatment of its effects may then be expected to be successful, and a detailed calculation of the properties of the effective electrons appears in principle feasible. If we compare our results with those of Kohn [17] for a single extra electron in an insulator, we see that our approach is somewhat more general (being not confined to a single electron), but somewhat less rigorous, since we have not proved the validity of the approximations I described. I believe that it offers a promising way to generalize

the results of KOHN to the case of many electrons in the conduction band, and that the generalization may be carried out along the lines I have described.

Note added in proof.

In a paper recently submitted to the *Phys. Rev.* NOZIÈRES and the writer have shown the following:

1) The calculation of the correlation energy by Gell-Mann and Brueckner is only valid for $r_s \lesssim 1$.

2) For actual metallic densities the contribution to the correlation energy from momentum transfers $k \lesssim 0.47 r_s^{\frac{1}{2}} k_0$ may be calculated accurately within the random phase approximation. This may be done by using the extended Hamiltonian (15), the canonical transformation of Sect. **4**, and summing all the higher order terms arising from $H_{\text{r.p.}}$ within the random phase approximation. The result through terms of order β^5 is

$$E_{\text{corr}}^{\text{l.r.}} = -0.458 \frac{\beta^2}{r_s} + 0.866 \frac{\beta^3}{r_s^{\frac{3}{2}}} - 0.98 \frac{\beta^4}{r_s^2} + 0.019 \frac{\beta^4}{r_s} + 0.706 \frac{\beta^5}{r_s^{\frac{5}{2}}} + \ldots \text{ Ry} .$$

The change in the β^4 contribution from the sum of (22) and (23) is due to the higher order terms from $H_{\text{r.p.}}$. β may be taken as $0.47 r_s^{\frac{1}{2}}$ and therefore $E_{\text{corr}}^{\text{l.r.}} \cong 0.043$ Ry.

3) Within the accuracy of a perturbation theory calculation, the contribution to the correlation energy from $H_{\text{s.r.}}$ arises only from electrons of anti-parallel spin, and is therefore given by (27).

4) The total correlation energy may be represented, to 15% accuracy, by the formula:

$$E_{\text{corr}} = -0.115 + 0.031 \ln r_s \text{ Ry} .$$

5) The foregoing result is very nearly the same as that obtained recently by HUBBARD (*Proc. Roy. Soc.*, A **243**, 336 (1958)).

6) The long range correlation energy may also been obtained by a suitable expansion of the result of Gell-Mann and Brueckner. The result agrees with that obtained using the BP approach and furnishes an explicit justification of the BP [1] neglect of the subsidiary conditions in their calculation of the ground state energy.

REFERENCES

[1] D. PINES and D. BOHM: *Phys. Rev.*, 85, 338 (1952), hereafter referred to as BP-II; D. BOHM and D. PINES: *Phys. Rev.*, **92**, 605 (1953), hereafter referred to as BP-III; D. PINES: *Phys. Rev.*, **92**, 626 (1953), hereafter referred to as P-IV.

A review article, which summarizes and extends the work described in these papers is D. PINES: *Solid State Physics*, Vol. **1** (ed. by F. SEITZ and D. TURNBULL), p. 367 (Academic Press, 1955), hereafter referred to as SSP.

[2] P. NOZIÈRES and D. PINES: *Phys. Rev.*, to appear in the Feb. 1, 1958 issue of the *Phys. Rev.*, hereafter referred to as NP-I, NP-II and NP-III.

[3] M. GELL-MANN and K. A. BRUECKNER: *Phys. Rev.*, **106**, 364 (1957).

[4] E. P. WIGNER: *Trans. Faraday Soc.*, **34**, 678 (1938).

[5] For a detailed discussion of the role of the subsidiary conditions, see D. BOHM, K. HUANG and D. PINES: *Phys. Rev.* **107**, 71 (1957).

[6] E. P. WIGNER: *Phys. Rev.*, **50**, 1004 (1934).

[7] O. KRISEMENT: *Phil. Mag.*, **2**, 245 (1957).

[8] N. F. MOTT: *Proceedings of the X Solvay Congress*, Bruxelles (1954).

[9] G. RUTHEMANN: *Naturw.*, **29**, 648 (1941); **30**, 145 (1492); *Ann. der Phys.*, (6), **2**, 113 (1948).

[10] W. LANG: *Optik*, **3**, 233 (1948).

[11] For a review of the experimental technique and the results up to 1955 in this field, see L. MARTON, L. B. LEDER and H. MENDLOWITZ: *Advances in Electronics and Electron Physics*, Vol. **7** (New York, 1955).

[12] See for instance, paper II of reference [2].

[13] D. PINES: *Rev. Mod. Phys.*, **28**, 184 (1956).

[14] M. GELL-MANN: *Phys. Rev.*, **106**, 369 (1957).

[15] L. D. LANDAU: *Journ. Exp. Theor. Phys. USSR*, **30**, 1058 (1956); English translation in *Sov. Phys.*, **3**, 920 (1957).

[16] See SSP and a forthcoming paper by J. FLETCHER and D. LARSON: *Phys. Rev.* (to be published).

[17] W. KOHN: *Phys. Rev.*, **105**, 509 (1957).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Transport Properties of Solids.

J. M. ZIMAN

Cavendish Laboratory, University of Cambridge - Cambridge

1. – Conduction by phonons.

1·1. *Perfect crystal.* – The Debye theory of specific heats leads directly to a theory of the conduction of heat in solids. The lattice waves may not only stand, and store thermal energy; they may also travel and carry energy. There is no difficulty in understanding this. The problem is more to understand why, in a perfect crystal, the conductivity is not infinite as if the lattice waves were like radiation in an empty cavity.

First note some properties of lattice waves. They are characterized by a wave vector, $\boldsymbol{q}$, which may take any value inside the first Brillouin zone, and by a frequency ν which is a function of $\boldsymbol{q}$. The energy in the wave is quantized in units of $h\nu$. We say that we have a certain number of « phonons » in the mode $\boldsymbol{q}$, just as we refer to light quanta as photons. The energy is carried with the group velocity $v = \mathrm{d}\nu/\mathrm{d}q$.

Generally speaking the phonon spectrum breaks into two parts, corresponding to the polarization of the lattice displacement in the wave. The longitudinal modes, where the displacement is more or less parallel to the wave vector, travel rather faster, and have a higher frequency, than the transverse modes. This separation is only approximate, and should not be relied on exactly. For long waves (small q) it is usually correct to take $\nu = vq$, with v constant; but, because of crystal anisotropy, the value of v may depend strongly on the direction of $\boldsymbol{q}$.

Thermal resistance in pure perfect crystals arises from the scattering of one phonon by another. The origin of this scattering is the set of anharmonic terms in the expansion of the lattice potential energy as a function of the displacements.

Thus, symbolically,

$$V = V_0 + V_2^{ij} y_i y_j + V_3^{ijk} y_i y_j y_k + \cdots ,$$

where y_i are displacements at sites i in the lattice etc.. We usually only go as far as the second term when we calculate the lattice frequencies. One can show that the next terms give rise to processes in which, say, two phonons collide, destroy one another, but create a third. These processes are, of course, subject to energy conservation conditions,

$$\hbar\nu + \hbar\nu' = \hbar\nu'' .$$

There are also interference conditions from the fact that each wave has a space factor like $\exp[i\boldsymbol{q}\cdot\boldsymbol{r}]$. These conditions are

$$\boldsymbol{q} + \boldsymbol{q}' = \boldsymbol{q}'' + \boldsymbol{g} ,$$

where $\boldsymbol{g}$ is a vector of the reciprocal lattice. If $\boldsymbol{g}$ were zero, this would look like a rule for the conservation of momentum.

We often think of $\hbar\boldsymbol{q}$ as the momentum of the phonon, but it is not really momentum at all, it is only a formal analogue. This is clear when we have $\boldsymbol{g}$ non-zero, for then, we should be saying that the (quasi) momentum is being created or destroyed in the scattering. The reason why we can have Umklapp processes (as we call it, when $g \neq 0$) is that $\boldsymbol{q}$ is not strictly defined more closely than by the addition of a reciprocal lattice vector (just as in the Bloch theory of electrons).

It is important to distinguish between N-processes ($g = 0$) and U-processes ($g \neq 0$), because the former do not give rise to a thermal resistivity. The gas of phonons one might say, is created at the hot end of the crystal and flows down, as if through a tube, to the cold end. N-processes do not hinder this flow; they are like the collisions between particles in a flowing gas, which give rise only to a viscosity, and this can only slow the flow if the walls are taken into account. U-processes on the other hand, can make the gas give up some of its momentum, and do cause the gas to be slowed, i.e. resist the diffusion of heat.

The full theory of this effect is exceedingly complicated. The main argument was given by PEIERLS [1]. At high temperatures ($T > \Theta$) scattering by U-processes, roughly proportional to the density of scattering phonons, is proportional to the magnitude of the fluctuations of density in the solid, which by thermodynamics is proportional to temperature. Thus $K \propto 1/T$. Since U-processes can only occur if the three phonon wave vectors, $\boldsymbol{q}$, $\boldsymbol{q}'$ and $\boldsymbol{q}''$ are all of the order of magnitude of $\boldsymbol{g}$ which is about twice the radius of the Debye sphere in reciprocal space, there is a strong tendency for these to be « frozen out » at low temperatures. Roughly speaking, the probability of a U-process is proportional to $\exp[-\Theta/2T]$. The thermal conductivity then

increases very rapidly when $T \ll \Theta$, i.e.

$$K \propto T^r \exp[\Theta/bT] \qquad (r \simeq 2,\ b \simeq 2).$$

If we accept this explanation, and ignore certain sophisticated details, we expect the thermal conductivity of a perfect crystal to rise as shown in Fig. 1. This sort of curve can be observed, and the theory is well verified in principle, if not yet made fully qualitative. But real solids are often not pure or perfect, and it is interesting to try to study the effects of various sorts of imperfections on the thermal conductivity. This is the topic we shall here pursue, since the theory for perfect crystals is algebrically complicated, not very complete, and best read out of the book by PEIERLS [2].

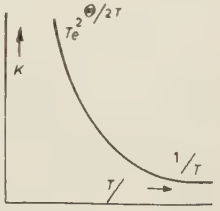


Fig. 1. – Typical thermal conductivity of an insulator.

Various crystal imperfections and impurities may be expected to scatter phonons. The scattering of a given phonon may be expected to depend very much on its wavelength, since this can vary enormously over a range from one to a thousand lattice spacings. The overall thermal conductivity will depend on some sort of average of the scattering of phonons of all wavelengths, and this average will depend on the distribution of phonons in various lattice modes, which, in turn, depend on the temperature. Thus, one must find some rule for this average.

Here we often use a general formula based upon kinetic arguments

$$K = \tfrac{1}{3} c v l \,,$$

where c is the lattice specific heat, v the group-velocity and l the mean free path of a phonon. This formula is crude but at least serves to separate out the significant physical quantities. If l and v were the same for all phonons, this formula would be true.

Now suppose we put into the crystal imperfections that scatter phonons of wavevector $\boldsymbol{q}$ at a rate which depends on $\boldsymbol{q}$.

We could from first principles calculate $l(q)$, say, a mean free path for phonons of that wavelength, which we should put for l in the kinetic formula. There are a number of alternative arguments, none exact.

a) Dominant mode. At the temperature T. There will be some frequency ν, for which the number of phonons is greatest. Most of the heat current will be carried by phonons of frequency $\bar{\nu}$.

Roughly speaking $h\nu \simeq kT$. This gives the frequency of an average phonon

at that temperature $\bar{\nu} \simeq kT/h$. This would give

$$q \simeq \frac{kT}{hv} .$$

If, as often happens, $l(q) \propto q^r$, then, also $l \propto T^r$. At low temperatures, when $c \propto T^3$, we get the useful empirical rule $k \propto T^{3+r}$.

b) Independent modes. One might try to be more sophisticated and write

$$K = \frac{1}{3}\int c(q)\, v(q)\, l(q)\, \mathrm{d}q ,$$

as the sum of the contributions of the various modes to the total conductivities, each contribution being a replica of the kinetic formula. This formula, also, gives the same power law rule, but unfortunately the integral often does not converge. Thus, if we remember that $c(q)$ would be proportional to q^2, i.e. to the number of modes of that vector, we get:

$$K \propto \frac{1}{3}\int_0^{q_{\max}} q^2 q^r \, \mathrm{d}q ,$$

which diverges at $q = 0$ if $r \leqslant -3$. In other words, the long waves are scattered so little that they carry all the heat.

c) Strong N-processes. The use of independent modes ignores the tendency of phonons to interchange into each other's modes, by N-processes (even if there are no U-processes to give resistance). If this mixing is very strong, one cannot just add the contribution of seaprate modes. One can show (Ziman [3]), as a limiting case, that one should then take

$$\frac{1}{l} = \frac{1}{cv}\int \frac{c(q)\, v(q)}{l(q)}\, \mathrm{d}q ,$$

as the average value of the inverse free path. This gives, in effect, an average of the probability of scattering by the imperfections, over all wavelengths, rather than an average of the m.f.p. itself. The integral always converges, and gives the usual power law at low temperatures. One can show that this formula gives a lower bound to the possible thermal conductivity of the solid.

d) Klemens formula. This avoids the singularity at $q = 0$ in the method of independent modes by a cut-off; $l(q)$ is taken to be constant for all values of q less than $\bar{q}$ where $\bar{q} = kT/\hbar v$. This method (Klemens [4-5]) is

based on a study of the properties of N-processes. It gives conductivities greater than *c*). Also thermal resistivities from different sources (from point defects or dislocations) are not additive. That is, Matthiessen's rule is not obeyed. For practical purposes, and first guesses, method *a*) is adequate. One says, roughly speaking, that the wavelength of the dominant mode is about Θ/T lattice spacings at the temperature T, but there may be quite a large numerical correction (say $(2\pi)^r$) when it comes to an exact formula.

1·2. *Effects of imperfections.* – Now we study the various imperfections, according to their dimensionality (see KLEMENS [6]).

(0) *Point imperfections.* These may be interstitial atoms, impurities, isotopes, or vacancies. The scattering is given by an old classical theory, which depends chiefly on geometry. All we need to do is to match incoming and outgoing waves on the surface of a small sphere about the imperfection. If this sphere has radius a, then the Rayleigh scattering law for the scattering cross-section is

$$\sigma \propto \left(\frac{a}{\lambda}\right)^4 ,$$

whence

$$l(q) \propto \left(\frac{1}{aq}\right)^4 , \qquad \text{i.e.} \quad r = -4 .$$

In principle, we can distinguish between 3 effects.

i) Mass. An extra mass, δM at a point in a medium of average density ϱ gives a scattering cross-section

$$\sigma = \frac{q^4}{4\pi\varrho^2}(\delta M)^2 .$$

This is an exact formula. Note that the presumed radius a of the sphere does not enter, although, if we took $\varrho = M/a^3$ so that a was the « atomic radius » we should get a formula of the type suggested,

$$\sigma = \frac{a^2}{4\pi}\left(\frac{\delta M}{M}\right)^2\left(\frac{2\pi a}{\lambda}\right)^4 .$$

For isotopes $\delta M/M$ may be small, perhaps a few percent. But it must be remembered that an isotope may be present as an « impurity » in great abundance. One can have (as in chlorine) a mixture in nearly equal proportions of two different isotopes. When one comes to calculate the mean free path, or the thermal conductivity, the scattering cross-section must be multiplied by this abundance. In fact, we now believe that the thermal con-

ductivity of most solids is dominated by the effect of isotopes just in the range where we should be seeing U-processes being frozen out.

The power law for this case would give

$$K \propto 1/T ,$$

right down to quite low temperature, instead of $K \propto \exp[\Theta/2T]$. For vacancies and interstitials this will also be a large effect, only there are also

ii) Elastic forces. The forces on an interstitial atom change the local elastic properties of the medium. Classical theory would replace $(\delta M/M)^2$ by $3(\delta\mu/\mu)^2$, where $\delta\mu$ is the charge in the local modulus of rigidity, or bond strengths or some similar parameter. The difficulty is to guess the amount of this.

iii) Strain anharmonicity. A vacancy or interstitial causes a strain in the surrounding lattice. The strained material has different elastic properties, due to the anharmonic terms in the elastic potential energy. The strength of these is measured, crudely, by the Grüneisen parameter γ, which appears in the theory of thermal expansion of solids. However, this effect is zero in an isotropic medium (KLEMENS [6]), and is otherwise very difficult to calculate.

(1) *Line imperfections.* – The typical one dimensional imperfection in a solid is a dislocation. The scattering effect of a dislocation for lattice waves can be thought of as divided into two portions. The difficulty is that the strain field spreads out a long way, on a scale much larger than the wave length, λ of the phonon.

i) Dislocation core. For the region of radius $r < \lambda$ we have Rayleigh scattering by a cylinder. This goes with a cross section (per unit length of dislocation) $\sigma \propto (a/\lambda)^3$, where a is, say, an effective radius of the core of the dislocation where the lattice disturbance is large.

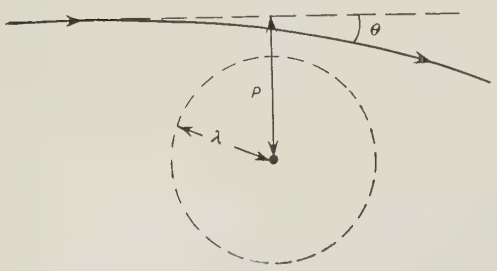


Fig. 2. – Deviation of an electron near a dislocation.

Here we should also include the scattering by the local strain field, just as for point defects. Thus the scattering cross-section would include further factors like $\gamma^2(\Delta V/V)^2$ for the dilatation effect.

ii) Main strain field. For the region $r > \lambda$, the Rayleigh theory will not apply. Now we are in the region of geometrical optics. The beam

passing at a distance p from the dislocation is in a region where the strain is of the order of b/p, where b is the Burgers vector. The angle of refraction comes out at $\theta \sim \gamma(b/p)$. If this is a small angle, the total cross-section will be proportional to

$$\int (1 - \cos\theta)\, \mathrm{d}\theta \simeq \int \left(\gamma \frac{b}{p}\right)^2 \mathrm{d}p \simeq \frac{\gamma^2 b^2}{p_0},$$

where p_0 is the least value of p. On our hypothesis, geometrical optics only works down to a radius $p_0 \sim \lambda$. Thus

$$\sigma \simeq \frac{\gamma^2 b^2}{\lambda} \qquad \text{and} \qquad l \propto \frac{1}{q}.$$

This is actually the main effect, since the wavelength of a typical phonon is much larger than the « radius of the core of a dislocation », or than a Burger's vector. These results have been derived by Klemens using the Born approximation but they are not specifically quantal processes. Thermal resistivities varying as $1/T^2$ have been observed in certain alloys and crystals where a high density of dislocations would be expected.

iii) Flutter. There is a further possible mechanism for scattering of lattice waves by dislocations. We have treated the dislocation and its strain field as if it were fixed to the medium. We know that, in fact, dislocations move very freely in metals; there is almost negligible resistance to motion of a dislocation along its slip plane. Now a shear wave incident on a screw dislocation has a strange effect. As pointed out by NABARRO [7], it causes a motion of the whole dislocation normal to both the direction of propagation of the wave and the dislocation line. The shear wave sets up a local strain field, to which the dislocation responds by slipping.

Crudely speaking, the dislocation moves for the time of one period before reversing. The distance moved is thus proportional to $1/\nu$, so is proportional to λ. In its motion the dislocation sends out other waves in all directions, so that, in sum, the incident phonon is scattered. The cross-section for this scattering turns out to be proportional to the range of the displacement of the dislocation $\sigma \propto \lambda$ or $l \propto q$. This also is a classical result: a rigid cylinder free to move in a fluid of the same density scatters sound waves in a similar way. In greater detail we have to include the « inertia » of the dislocation, which is proportional to $\ln \lambda/b$. There is also the Peierls-Nabarro force keeping the dislocation to a special lattice position. Perhaps the dislocation only moves in the bottom of a trough in the lattice field. This would explain why there is no effect of this sort in covalent crystals, where the dislocations are tightly bound by molecular bonds.

Evidence for this effect (giving $K \propto T^4$ at low temperature) is meagre, but it might explain some anomalous effects in some superconducting metals.

(2) *Surfaces of imperfections.* These are more complicated and we know very little about their scattering power. Stacking faults and twin boundaries, for example, probably have very little effect on lattice waves, except those of very short wavelength. The imperfection of the lattice is confined to a thin layer and the medium has nearly the same properties on both sides.

Grain boundaries show two scattering mechanisms.

i) Acoustic mismatch. In a typical crystal the velocity of sound is anisotropic, with, at most, cubic symmetry. The propagation of elastic waves depends on direction. At a tilt boundary there will be a mismatch of velocities on the two sides; and consequently partial reflection. The theory is again classical, as for seismic waves, and depends on the polarization in a complicated way. The main result is that geometrical optics should apply; the reflection coefficient would be independent of wavelength giving a thermal conductivity proportional to T^3. The scattering probably also depends on the square of the angle of mismatch.

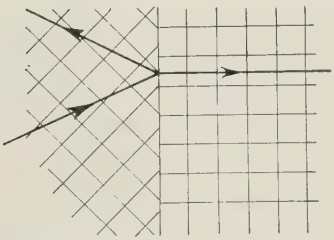


Fig. 3. – Acoustic mismatch.

ii) Dislocation arrays. Grain boundaries always tend to break into arrays of dislocations, especially at small angles. The boundary region then acquires a certain thickness, which is, roughly speaking, the diameter, $2a$, of the cylindrical core of a single dislocation. The array can give Rayleigh scattering with $\sigma \propto (2a/\lambda)^2$. This is probably not a big effect.

Crystal boundaries are a special form of two dimensional imperfections. That is, we suppose, our thermal conductivity to be measured in an infinite crystal and are distressed to have to use specimens only a few millimeters in diameter. As we have seen the U-processes scattering decreases rapidly as we go to low temperature, so that the m.f.p. may easily become comparable with the crystal size. This case was early calculated by Casimir [8] but goes back to the problem of Knudsen gas solved by Smoluchowski.

i) Rough walls. In the simplest case where the scattered phonon does not remember its direction, one can show that l is equal to the diameter D, of the specimen, or at least some average of the same sort if it is not of circular cross-section. This makes

$$K \propto T^3 ,$$

which is the typical form of the conductivity at low temperature. But note

that this « conductivity » is not absolute but is proportional to the size of the specimen.

ii) Smooth walls. Suppose that the crystal surfaces are rather smooth. For the sake of calculation, suppose that a fraction p of the phonons are specularly reflected. Then one can show that

$$l = D\left(\frac{1+p}{1-p}\right),$$

which can become rather bigger than D. This is observed. It is not so easy to calculate p, for this should vary with phonon wavelength. Suppose we take A to be the average (root mean square) variation of surface height. Then one would expect the following limiting behaviour: $\lambda \ll A$, $p = 0$; the waves come off with random phase and are scattered irregularly: $\lambda \gg A$, $p = 1$; the wavelength is too great even to see the asperities. Crude calculation suggests that the transition from one type of scattering to another is quite sharp so that one could reasonably take $p = 0$ for $\lambda < A$ and $p = 1$ for $\lambda > A$. The theory of scattering of radio waves from the monosphere is very helpful on this point. It seems, however, that one cannot just take a definite value of A, and see l increasing suddenly as the dominant phonon wavelength passes

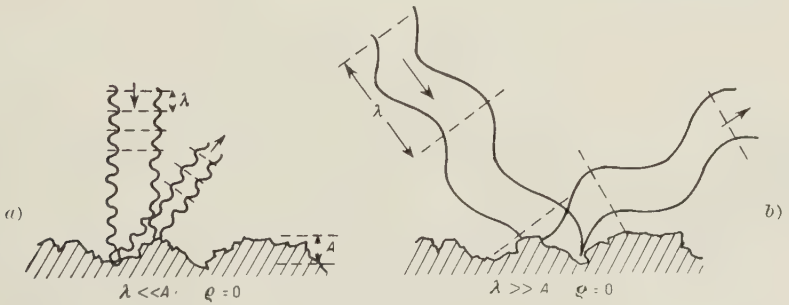


Fig. 4. – *a*) Short waves see a rough surface; *b*) Long waves see a smooth surface.

through this point. Crystal surfaces are not just random functions. It is better to think of them as made up of a patchwork of areas; on the fraction p of the surface area, the value of A is less than λ.

This gives a reasonable picture of the scratched, stepped, badly polished surface.

Fig. 5. – Typical crystal surface.

iii) Saw tooth. Suppose that the crystal surface has a definite zig-zag pattern cut into it.

A phonon trying to travel down the specimen will be preferentially reflected back the way it came. This can reduce the thermal conductivity making

$$l < D \, .$$

Fig. 6. – Scattering by saw-tooth wall.

An effect of this sort seems to have been observed by WHITWORTH, in liquid helium in a tube with large irregular lumps on the walls.

(3) *Three-dimensional imperfections.* In a glass one cannot locate the imperfection of the crystal lattice at special points, lines or surfaces. The material is irregular all through. There is another case where the theory of propagation of radiowaves in the ionosphere is helpful. One needs only two parameters to define the problem: $(\Delta v)^2/\bar{v}^2$, which is the relative variance of the velocity of a wave from point to point in the medium and L the correlation length, which measures the scale of distances over which the refractive index may be supposed constant. We find then, two limits:

When $\lambda \ll L$, the wavelength of the waves is smaller than the regions into which the volume is dissected by L. Passing from one region to the next the waves are scattered with a m.f.p. proportional to L,

$$l \simeq \frac{\bar{v}^2}{(\Delta v)^2} L = \text{constant} \, .$$

But, when $\lambda \gg L$, the long waves ride over the varying refractive index, only acquiring a certain amout of randomness of phase by the addition of small effects from each region. One then finds

$$l \sim \frac{\bar{v}^2}{(\Delta v)^2} \frac{\lambda^2}{L} \propto \frac{1}{q^2} \, .$$

This is verified for glasses. Taking K to be about 10 Å.U. the size of a typical ring of silicate groups, we get a constant thermal conductivity at high temperature, where λ is of the order of the lattice spacing, but at lower temperatures the conductivity is proportional to T.

2. – Electronic conduction in metals.

Although a perfect crystal of a solid with a high Debye temperature (e.g. Diamond) is actually a very good heat conductor, we usually think of typical conductors being metals. That is, we think mainly of transport by the electrons, and specifically of electrical conductivity. The complete theory

of electron transport in solids is very complex algebrically, especially when we think of the effects of both electrical and thermal gradients, as in the thermo-electric phenomena; or when we add a magnetic field and twist up the electron orbits. I shall only try to give an account of the microscopic processes which give rise to electrical resistivity in metals, concentrating on the general physical picture rather than on the mathematical details. We know that the electrons in a metal are free to move subject to certain conditions arising from the exclusion principle.

We know that the electron states may be classified according to their quasimomentum $\hbar\boldsymbol{k}$, with only two electrons in each state. We know that most such states lie deep in energy, and are always filled, whilst others are always empty. We know that one can define the Fermi surface, a surface of constant energy in $\boldsymbol{k}$ space, such that, to a good approximation all states within that surface are filled and all outside it are empty.

The effect of an electric field $\boldsymbol{E}$ is to accelerate all the electrons. The force $e\boldsymbol{E}$ is just the rate of change of quasimomentum, i.e.

$$e\boldsymbol{E} = \hbar\dot{\boldsymbol{k}} .$$

Thus the Fermi surface, and all that is therein, is displaced from its equilibrium position at a uniform rate. Where does it stop? There must be other processes opposing the displacement. These are the relaxation or scattering processes, which take an electron at A (going rather faster than usual) and send it to B, where there is a deficiency of fast electrons.

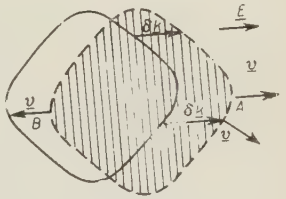


Fig. 7. – Displacement of the Fermi surface.

For the moment, we sum up these relaxation processes in a relaxation time τ defined by a typical exponential decay equation, e.g.

$$\dot{\boldsymbol{k}} = (\boldsymbol{k}_F - \boldsymbol{k})/\tau ,$$

where $\boldsymbol{k}_F$ is value of $\boldsymbol{k}$ in the absence of a field. If these two effects balance, we have

$$\delta\boldsymbol{k} = (\boldsymbol{k} - K_F) = \frac{\tau e\boldsymbol{E}}{\hbar} ,$$

but this extra corresponds to a net electron current, $\boldsymbol{J}$.

Thus

$$\boldsymbol{J} = \frac{e}{4\pi^3}\int \boldsymbol{v}\,\mathrm{d}\boldsymbol{S}\cdot\delta\boldsymbol{k} ,$$

since $\mathrm{d}\boldsymbol{S}\cdot\delta\boldsymbol{k}$ is the extra volume in $\boldsymbol{k}$ space where there is a density $1/4\pi^3$ of electron states.

Substituting for $\delta\boldsymbol{k}$ we have:

$$\boldsymbol{J} = \frac{e^2}{4\pi^3\hbar}\tau\int \boldsymbol{v}\,\mathrm{d}\boldsymbol{S}\cdot\boldsymbol{E} = \boldsymbol{\sigma}\cdot E\,,$$

where $\boldsymbol{\sigma}$ is the electrical conductivity tensor. Taking account of averages over direction, which allows us to write, for a cubic crystal

$$\int \boldsymbol{v}\,\mathrm{d}\boldsymbol{S} = \frac{1}{3}vS\,,$$

(with v the average Fermi velocity) the formula for the conductivity is

$$\sigma = \frac{e^2 S}{12\pi^3\hbar}\,\Lambda\,,$$

(with $\Lambda = \tau v =$ the mean free path of an electron).

A more sophisticated derivation, using the Boltzmann equation, verifies this formula, and also defines Λ and τ.

This is given by

$$\frac{1}{\tau} = \int Q(\theta)(1-\cos\theta)\,\mathrm{d}\Omega\,,$$

where $Q(\theta)$ is the probability of an electron being scattered into the solid angle $\mathrm{d}\Omega$, making an angle θ with its original direction. This formula applies, of course, only to spherically symmetric scatterers such as impurity atoms.

The main point to be emphasized is that the conductivity is proportional to the area S, of the Fermi surface, and to mean free path, Λ, of the electrons. This is why metals with large areas of Fermi surface (Alkali metals, Noble metals, Aluminium, etc.) are good conductors in general. The values of S is a quantity to be determined by other means, or from calculation. We shall not consider it further here, but will study the behaviour of Λ for various types of scattering objects.

1) *Point imperfections.* The theory of the scattering of electrons by isolated impurities in metals is given by FRIEDEL, and will not be discussed here.

2) *Dislocations.* One thinks of a dislocation as a narrow disturbed core surrounded by a widespread strain field. The latter has the main effect in scattering electrons. Here the method of deformation potentials is extremely valuable.

We take an element of volume some distance from the axis. In this element the strain is nearly homogeneous, and so we can calculate the energy of the

electronic state, the function ε_k, as if this element were part of a large lump of the same sort of material. If the strain is η, we could treat

$$\varepsilon_k(n) - \varepsilon_k(0) \simeq \eta \frac{\partial \varepsilon_k}{\partial \eta},$$

as the deformation potential acting on an electron of wave vector $\boldsymbol{k}$, scattering it. As η varies slowly over the strain field round the dislocation, we could use the Born approximation to calculate the cross-section.

But this argument needs some amplification. The fact that our system of ions is immersed in an electron gas means that we must allow for the reaction of the « outer » electrons, which can serve to screen out this sort of deformation potential, just as they could screen out the charge of an impurity. In a shear strain this is not important, because the extra electron density gives the same change in the potential energy of all electrons moving in different directions relative to one another. That is, the constant energy surfaces in $\boldsymbol{k}$-space are sheared. As the direction of shear will be different in different parts of the strain field of the dislocation, an electron of given $\boldsymbol{k}$ vector has different energies, and is scattered.

In a dilatation, however, we have to take into account the change in density of the ions. If there are n ions per unit volume, a dilatation Δ will give rise to a charge density $(ne\Delta)$.

Electrons to nearly this number will be drawn in so that the Fermi level would rise by the amount

$$n\Delta \Big/ \frac{\partial n_0}{\partial \varepsilon} = \delta \varepsilon_F .$$

But this cannot be allowed, by the Thomas-Fermi principle. Just enough charge unbalance will be permitted so that there will be an electric field, with

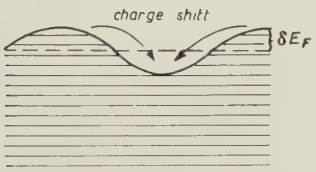


Fig. 8. – Dilatation of a crystal: *a*) before charge shift.

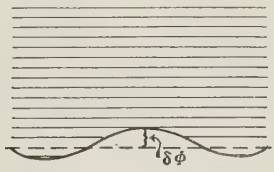


Fig. 9. – *b*) after charge shift.

electrostatic potential $\delta\varphi$, which can be subtracted from all the electron energies to bring the Fermi-level back to constant. This potential can scatter electrons, and its amount is just $\delta\varepsilon_F$. If the distance scale of the strain field is large, the number of electrons required to move to set up this potential is negligible.

For free electrons $(n/(\partial n_0/\partial \varepsilon)) = \frac{2}{3}\varepsilon_F$. Without charge shift the deformation potential would have been of about the same magnitude, but this is effectively screened out and replaced by this quantity. Thus, one can calculate the scattering by the dilatation in the dislocation strain field without further knowledge than the density of states.

There is a little difficulty in actually calculating the resistance, because one has to average over different directions of dislocations and so on. Thus, an array of dislocations with axes parallel would not offer any resistance at all.

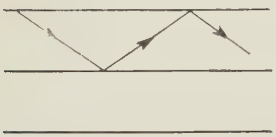


Fig. 10. — Scattering by dislocations.

The effect of the core of a dislocation is not easy to calculate explicitly. However, it is probably not serious, because the strain field goes as $1/r$ and the integrals calculating the scattering do not need to be cut off at any inner radius.

However, STEHLE and SEEGER [9] have emphasized the importance of the anharmonic elastic forces, which expand the core of a screw dislocation, and have calculated the extra resistance using a phase-shift method.

The formula for the extra resistivity of the whole dislocation is rather complicated, especially for edge dislocations where we have both shear and dilatation fields. According to HUNTER and NABARRO [10] reasonable results for the extra-resistivity would be about

	edge	screw	
for Cu	6	2	10^{-3} e.s.u. per dislocation line
Na	23	2	per sq. cm.

I should be prepared to bet that the results for Cu should be increased both for edges and screws, because the Fermi surface touches the zone boundary and is thus rather susceptible to shear strains.

3) *Grain boundaries* would scatter electrons by the same mechanism as they scatter phonons, i.e. through the mismatch of electron velocity when the Fermi surface is not spherical and through the presence of microscopically disturbed regions where the boundary has condensed into an array of dislocations. The short wavelength of electrons would make the latter a more important effect than it is for phonons.

The most interesting two dimensional defect in this respect is a stacking fault (*).

How should stacking faults scatter electrons? The first estimate is—not at all—. Free electrons would not be aware of the change of phase in the

(*) For a geometrical description reference is made to Prof. FRANK's lectures.

lattice at this boundary. There is no change of density. The general electric potential of an electron would be scarcely altered in the boundary layer because the number of nearest neighbours has not changed. One can say that a region of H.C.P. lattice has been interposed in the F.C.C. structure, but this is only a narrow layer and probably not thick enough to develope its characteristic electronic structure.

The next estimate is based on the augmented plane wave method of calculating electronic band structure. This analyses the potential in a crystal into spherical regions surrounding each ion and empty spaces in between. Plane waves in the interstices are matched on the spherical boundaries to solutions of the Schrödinger equation for the ionic potential. Then we take some suitable surface at the stacking fault in the empty region and match wave functions on both sides of the fault. Now if the A.P.W. consisted of just one plane wave ($\exp[i\boldsymbol{k}\cdot\boldsymbol{r}]$) in the interstices, the matching could be made perfect. Thus there would be no reflection—essentially we should still be treating free electrons.

But there will also be some admixture of higher terms $A_g \exp[i(\boldsymbol{k}+\boldsymbol{g})\cdot\boldsymbol{r}]$, where $\boldsymbol{g}$ is a reciprocal lattice vector. These waves are tied in phase to the lattice and therefore will not match exactly at the boundary. It turns out that we get a reflection coefficient proportional to $(A_g)^2$. Thus our next estimate is small (say 10^{-2}) because a simple plane gives good solutions of the band calculation in metals, so we do not expect A_g to be large.

However, more recent work shows that it may not be so. The experiments of Pippard on copper show that the Fermi surface touches the zone boundary in the (111) directions. Near these points the wave function must be a nearly equal mixture of $\exp[i\boldsymbol{k}\cdot\boldsymbol{r}]$ and $\exp[i(\boldsymbol{k}+\boldsymbol{g})\cdot\boldsymbol{r}]$ coming from an equivalent point in the extended zone scheme. Thus A_g is of the order $\frac{1}{2}$ and strong reflections would be observed. Since the stacking faults are normal to the [111] directions, the reflections would be very effective in resistance. As the experimental evidence seems to favour large reflection coefficients, this is fortunate. Note that the large resistance is bought at very little expense in energy in setting up the stacking fault. It is only a few electrons at the Fermi surface that are affected. Other electrons of longer wavelength pass through the wall without seeing it at all.

Twin boundaries are very similar to stacking faults and would be expected to have similar properties.

Crystal surfaces can be made to scatter electrons if we make very thin films or thin wires. The evidence is on the whole, that the surfaces behave as if rather rough with very little specular reflection. The problem of calculating in such cases as usually simply geometrical with the complexity that the mean free path of the electrons for other causes (residual resistance) may be very short compared with the thickness of the wire etc. (see Sondheimer [11]).

4) *Disordered structures* A solid containing an irregular distribution of mass, or of elastic forces, could scatter phonons. In the same way, an irregular distribution of electric forces will scatter electrons. We saw that long wave phonons could glide over the irregularities of mass, and hence would not be scattered so effectively. Electrons on the Fermi surface are all of rather short wavelength, so will have more or less constant cross-section.

But now we have to obtain an irregular distribution of potential. This comes most simply in a disordered alloy. We might have a proportion x of A atoms, with potential V_A, and a proportion $(1-x)$ of B atoms with potential V_B. Then the average potential seen by each electron will be

$$V = xV_A + (1-x)V_B .$$

At an A atom the deviation of the potential from the average is

$$V - V_A = (1-x)(V_B - V_A)$$

and at a B atom it is

$$V - V_B = x(V_A - V_B) .$$

The scattering probability is the square of the matrix element: for an A atom it is proportional to $(1-x)^2(V_A - V_B)^2$. But the density of A atoms is x. Thus, adding effects of A's and B's, we get:

$$\varrho \propto x(1-x)^2(V_A - V_B)^2 + (1-x)x^2(V_B - V_A)^2 \propto x(1-x)(V_A - V_B)^2 ,$$

which is a typical formula for the resistivity of a disordered alloy.

When the alloy becomes ordered this resistance may be lost altogether because the electrons are scattered coherently by the regular superlattice of A and B atoms, and this only means that the Fermi surface, Brillouin zone, energy surface etc., adjust themselves slightly and the alloy behaves as if it were a pure perfect crystal.

Is this sort of adding up of effects of separate atoms, incoherently scattering, valid? This is a difficult problem of diffraction theory. The answer seems to be that it will be correct so long as the scattering is not too strong. The electron mean free path must not be comparable with the lattice spacing or the electron wavelength.

Another form of disordered scattering has recently been studied. This is scattering of electrons by the spins of d-electrons in transition metals. The conduction electrons may be coupled to the d-electrons by an exchange interaction of the form

$$J\delta(\boldsymbol{R}_1 - \boldsymbol{R}_2)\boldsymbol{S}_1 \cdot \boldsymbol{S}_2$$

(where the δ function is put in to avoid deciding exactly the amount of localization of interaction). The interest of this scattering is that it will depend upon the degree of disorder in the spin system. Thus, high above the Curie or Néel temperature in a ferromagnet or antiferromagnet, the spin system is disordered, and the scattering is incoherent, and a large resistivity is possible. As the temperature falls, we may get alignement of the *d*-spins, and hence the scattering becomes coherent, and ceases to create resistance. Near the Curie point itself there may be a certain amount of local ordering, i.e. correlation of spins on neighbouring atoms, and this may even in principle enhance the scattering if it coincides with the electron wavelength.

This mechanism seems to explain resistance anomalies in the rare earth metals. However, it is not so easy to distinguish it experimentally from another effect, the *s-d* scattering in the transition metals. Here it is supposed that a large part of the electrical resistance is due to the electrons being scattered from the *s*-band (where they are good carriers of current, with a large Fermi surface area) into the *d*-band which has a high density of states ready to receive them. In this transition, electron spin is not altered. But then, if the *d*-band is aligned, so that all the spins point the same way, there will be no room for electrons of that spin to be scattered there from the *s*-band. Thus, roughly speaking, half the electrons will not suffer any *s-d* scattering, and will carry a lot more current. The electrical resistivity will then drop, in a manner similar to the figure for spin disordered scattering, at the Curie point.

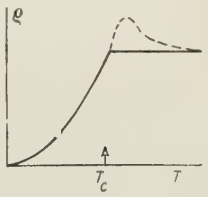


Fig. 11. – Spin-disorder resistance.

In principle, one could distinguish between these two mechanisms by the effect of adding impurities. Their contribution to resistivity above T_c for the *s-d* mechanism will double its low temperature value, whereas spin disordered resistance would show no change.

3. – Electron-phonon interaction.

We have treated electrons and phonons as if they were independent carriers of heat or electricity, and calculated their scattering by other objects in the crystal. In fact, they interact strongly with one another, and where both are present as in a metal this interaction is the dominant effect. For example, at low temperatures the conduction of heat by phonons is almost suppressed by the scattering of lattice waves by free electrons, so that the thermal resistivity of a metal is mainly due to conduction electrons. Conversely, the electrical resistivity of a pure metal, except at very low temperatures, is due mainly to scattering of electrons by the ordinary thermal lattice waves.

The theory of this scattering is complicated in detail, and still not fully known. The basic argument is that the displacement of the lattice, by an amount $\boldsymbol{y}$, say, changes the local electron potential by some amount δV, e.g.

$$\delta V = \boldsymbol{y} \cdot \boldsymbol{U} ,$$

where $\boldsymbol{U}$ is a quantity like ∇V. An electron will then be scattered with a matrix element like

$$\int \psi_{k'}^{*} \, \delta V \, \psi_k \, \mathrm{d}\boldsymbol{r} ,$$

from the state $\boldsymbol{k}$ to the state $\boldsymbol{k}'$. But if the displacement $\boldsymbol{y}$ is due to the presence of a lattice wave, e.g. if $y = b_q \exp [i\boldsymbol{q} \cdot \boldsymbol{r}]$ and if ψ_k etc., is of the Bloch form, we get a matrix element like

$$\left\langle b_q \int \exp [-i\boldsymbol{k}' \cdot \boldsymbol{r}] \exp [i\boldsymbol{q} \cdot \boldsymbol{r}] \, U(r) \exp [i\boldsymbol{k} \cdot \boldsymbol{r}] \, u_k(r) \, u_k^{*}(r) \, \mathrm{d}r \right\rangle ,$$

where the symbols $\langle \, \rangle$ mean « calculate the matrix element between phonon states ». Now in here the integral repeats itself from cell to cell, except for a factor like

$$\int \exp [i(-\boldsymbol{k}' + \boldsymbol{k} + \boldsymbol{q}) \cdot \boldsymbol{r}] \, \mathrm{d}\boldsymbol{r} .$$

which vanishes unless

$$\boldsymbol{k} + \boldsymbol{q} = \boldsymbol{k}' .$$

Also, the amplitude b_q of a harmonic oscillator has zero matrix element unless it is measured between two states differing by one quantum. Thus, the number of phonons in the system must change by unity. Again, we use time-dependent perturbation theory, which insists that the energy of the system is conserved in the process.

These conditions can all be reconciled if we think of our scattering process being described in particle terms as the collision of an electron (wave vector $\boldsymbol{k}$) with a phonon (wave vector $\boldsymbol{q}$) in which the latter is annihilated, and the electron scattered into a state $\boldsymbol{k}'$. In addition to the wave vector conservation condition, we have the condition

$$\varepsilon_{\boldsymbol{k}} + \hbar \nu_{\boldsymbol{q}} = \varepsilon_{k'}$$

for the conservation of energy.

The classical significance of these conditions is interesting. Suppose that $\boldsymbol{q}$ is small. Using the first condition we have

$$\varepsilon_{\boldsymbol{k}} + \hbar \nu_{\boldsymbol{q}} = \varepsilon_{\boldsymbol{k}+\boldsymbol{q}} \approx \varepsilon_{\boldsymbol{k}} + \boldsymbol{q} \cdot \nabla_k \varepsilon_k , \quad \text{i. e.} \quad \hbar \nu_{\boldsymbol{q}} \approx \boldsymbol{q} \cdot \hbar v_{\boldsymbol{k}} ,$$

where v_k is the group velocity of an electron in the state $\boldsymbol{k}$. Dividing by h, and remembering that ν_q/q is the velocity of sound, s, we say that the electron is scattered if $s = v_k \cos\theta$.

That is, the component of the electron velocity in the direction of motion of the wave equals the velocity of sound. An electron in a favourable position in the phonon field (in a « trough ») moves so as just to keep in phase with the wave. Thus, it keeps on drawing energy from the wave, and is scattered by it. This is the « surf-riding resonance ».

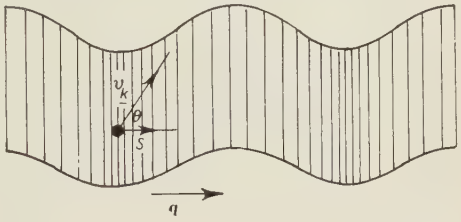


Fig. 12. – The surf-riding resonance.

But there are also non-classical effects to consider. For example, $\boldsymbol{k}$ and $\boldsymbol{k}'$ are always arbitrary, to the addition of a reciprocal lattice vector $\boldsymbol{g}$. The « momentum conservation condition » would then be perhaps.

$$\boldsymbol{k} + \boldsymbol{q} = \boldsymbol{k}' + \boldsymbol{g} .$$

We have electron-phonon Umklapp processes. These are very important. Think of a half full zone. The maximum value of $\boldsymbol{q}$ would be, say, the diagonal of this zone. This would be less than the diameter of the Fermi surface, so that $\boldsymbol{k}'$ would be unacessible to ordinary N-processes. But we can add on $\boldsymbol{g}$ (for example by repeating the whole zone as shown), and this can bring $\boldsymbol{k}'$ to easy reach of $\boldsymbol{k}$, so that scattering is possible.

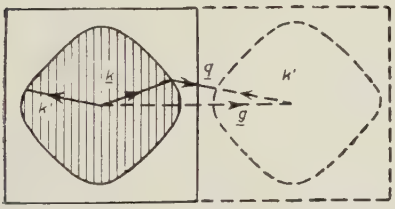


Fig. 13. – An electron-phonon U-process.

This scattering would be very important in electrical resistivity because it means an about-turn of the electron velocity. We still have to work out the cross-section of the scattering. This is a complicated and difficult problem, which has not been solved exactly.

The main results are as follows:

a) The interaction depends on the polarization of the lattice waves. Longitudinal phonons give the main scattering. It used to be thought that transverse phonons would not scatter electrons at all, but this is probably false. The polarization effect usually occurs as a factor $(\boldsymbol{k} - \boldsymbol{k}') \cdot \boldsymbol{r}_\alpha$, where $\boldsymbol{r}_\alpha$ is the polarization vector of the phonon. For N-processes, where $\boldsymbol{k} - \boldsymbol{k}' = \boldsymbol{q}$, this vanishes for transverse modes. For U-processes, however, this is not so.

One can also derive an interaction by considering the effect of shear strains on an electron. We saw that these could give rise to a deformation potential

in the scattering by screw dislocation. For the electron-phonon case, one only needs the Fermi surface not to be spherical to get N-processes by this mechanism.

b) The dominant factor in the interaction is an energy, often written C in the books. This derives mainly from the deformation potential due to the charge shift where the lattice is dilated by a longitudinal phonon. We saw for the case of dislocation scattering that this might have the value $\frac{2}{3}\varepsilon_F$, although there is some further correction if the phonon wavelength is short.

c) The cross-section varies strongly with angle of scatter. For N-processes it is more or less constant but it falls to small values at large scattering angles. This variation arises from integrals which depend on the form of $U(r)$ inside an atomic cell. Thus, it is important to know whether the ionic potential moves rigidly with the ion, or is deformed and smoothed as we go into the interstices of the lattice. The differential cross-section is rather sensitive to this and none of the formulae are exact. The smooth function given by BARDEEN [12] is probably the best bet, but there is no rigorous proof that some function looking more like the original BLOCH [13] formula, constant for N-processes and zero for U-processes, may not be more correct.

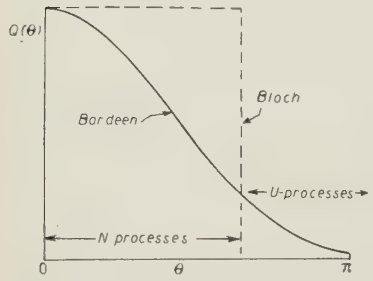


Fig. 14. – The electron-phonon scattering formula.

Transport properties. It is still a major calculation to derive formulae for the electrical and thermal conductivity. The usual method is to study the Boltzmann transport equation for the electrons, subject to collision with phonons. A good deal of work has gone into this and formally exact solutions can be developed (see, e.g. WILSON [14]). The general picture is as follows. For electrical conduction we only have to calculate a scattering relaxation time or mean free path, according to the formula

$$\frac{1}{\tau} = \int Q(\theta)(1 - \cos\theta)\,\mathrm{d}\Omega\,.$$

At high temperature, all the phonon modes are excited, so that $Q(\theta)$ will, for all θ, be proportional to the « density of lattice waves », i.e. by simple statistical mechanics, to kT. The integral over angle θ comes out as a constant. It is obvious that this nearly weights scattering through large angles. Thus, even though the scattering cross-section is small for U-processes, their net effect on the electrical resistance, is actually larger than for N-processes.

We have

$$\varrho \propto T$$

in this range.

At low temperatures many of the phonon modes are frozen into the ground state. Only those with small values of q are available to scatter electrons. Crudely speaking, only the « dominant phonons » those with $h\nu_q \approx kT$ need be considered. In each of these modes with $q = \bar{q}$, say, there will be a density of phonons proportional to T. But only a fraction of these, the ones correctly oriented to scatter the electron $\boldsymbol{k}$ and satisfying the conservation rules will be used. This fraction will depend on $\bar{q}^2$ which is proportional to T^2.

Again, for small angles of scatter, $(1 - \cos\theta) \sim \frac{1}{2}\theta^2$ which is also proportional to $\bar{q}^2$. Thus

$$\frac{1}{\tau} \propto T\bar{q}^2\theta^2 \propto T^5 .$$

The resistivity should thus be proportional to T^5. One can see that the resistivities here must depend strongly on the Debye temperature (as $(T/\Theta)^5$). If only longitudinal phonons scatter electrons this should be the value of Θ for transverse modes. The experimental results suggest that there is some admixture from the transverse modes, lowering the observed value of Θ. For N-processes, since only small scattering angles occur, the form factor of the scattering cross-section will not be important. The resistance will then depend on the limiting value at $\theta = 0$. From their geometry, there is a lower limit q_m to the value of q which may occur in U-processes. This is the shortest distance between the Fermi surface and its image in the repeated zone (see Fig. 13).

In principle, the resistance due to the U-processes should be frozen out at some low temperature, with perhaps an exponential factor $\exp[-(h\nu_{q_m}/kT)]$. In practice this is not observed. q_m happens to be a direction of low velocity transverse phonons (Brooks, private communication), so that $h\nu_{q_m}$ is a rather small fraction of $k\Theta$. Also the exponential factor is blurred by the variation of cross-section with angle and other geometrical rules. The effect is masked from observation by the residual resistance due to impurities at very low temperatures. A formula of the type

$$J\left(\frac{\Theta}{T}\right) = \left(\frac{T}{\Theta}\right)^5 \int_0^{\Theta/T} \frac{z^5\,\mathrm{d}z}{(e^z - 1)(1 - e^{-z})} ,$$

describes the resistivity quite accurately, but this is only an empirical fit and not an exact consequence of the theory.

Thermal conduction shows a different behaviour. At high temperatures, it

is true, we can use the Wiedemann-Franz law, asserting the constancy of the ratio $K/\sigma T$. This makes K come independent of T.

We can see this in another way by using the kinetic formula

$$K = \tfrac{1}{3} C_e v \Lambda_e$$

and taking for C the electronic specific heat, proportional to T, but using the m.f.p. Λ, proportional to $1/T$ for scattering by phonons.

At low temperatures this simple relation no longer holds. This is because the electrons can lose energy in scattering by phonons. Remember that a thermal conductivity is measured under conditions where there is no net electric current. Thus, we must have an excess of « hot » electrons travelling one way and an equal number of « cold » electrons travelling the other. To reduce the heat current one does not need to turn a hot electron round in the inverse direction, it has only to lose its heat, whilst still travelling in the same direction. But the excess of heat is about kT, and this can be lost by just one collision with a typical phonon. Thus in the scattering cross-section the factor $(1 - \cos\theta)$ is dropped. The effective relaxation time is just $\propto T^{-3}$. Combining this with the specific heat ($\propto T$) we get

$$K \propto T^{-2} .$$

The thermal conductivity rises much less steeply than the electrical conductivity. The exact form depends on the form of the scattering cross-section, on the geometry of the zone, etc. An important point is that U-processes contribute in a rather different way from N-processes. In the original Bloch formula it appears that there should be a minimum in the thermal conductivity. This is not observed. When we allow in the theory for the variation of cross-section, and also for the properties of U-processes, this minimum almost vanishes (Ziman [15]). This is, at present, the best evidence we have that U-processes really are important in electronic conduction. Notice that in principle the Wiedemann-Franz law breaks down entirely at low temperatures although in practice it is restored by the impurity scattering for which it is always obeyed.

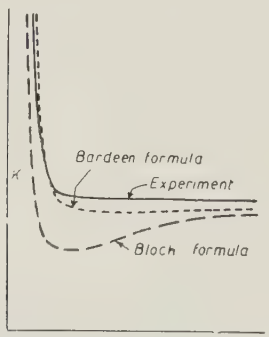


Fig. 15. – Thermal conductivity of a monovalent metal.

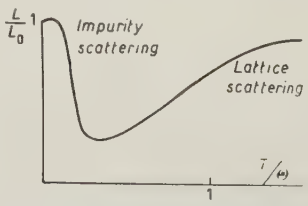


Fig. 16. – Wiedemann-Franz ratio (Lorenz number).

Thermoelectric effect. This really requires a lecture to itself. The only point to be made here is that if there is a thermal gradient there will be a current of phonons through the solid, as well as any electronic currents carrying heat. The phonon current can, so to speak, sweep electrons with it, and hence enhances the Seebeck potential required to restrain the electrons. At low temperatures this can greatly increase the thermoelectric power in alkali metals and, as mentioned by Professor FAN, in semiconductors.

It is very important not to assume crudely that the phonon distribution is in equilibrium when calculating the scattering of electron in such a case. It is not certain whether there is observable an inverse effect whereby the phonons are dragged by the electrons in electrical conduction, and hence fail to scatter quite so efficiently.

Lattice conduction. The electrons scatter phonons. Roughly speaking, the probability of scattering is just proportional to the number of electrons that are available for scattering; that is the number in the thermal layer on the Fermi surface. This is proportional to T, so that the phonon m.f.p. is $\propto 1/T$, and the lattice component of thermal conductivity is then proportional to T^2 at low temperatures. This can be observed by adding impurities to the metal until the electron m.f.p. is so small that the electronic heat conduction does not outweight the lattice conduction.

There is an important additional effect. Suppose we add so many impurities that Λ_e is much less than the wavelength of a typical phonon, i.e. $q\Lambda_e < 1$. The model of independent excitations interacting with a certain matrix element

$$\left\langle \int \psi_{k'}^{*} \{ b_q u(r) \exp [i\boldsymbol{q} \cdot \boldsymbol{r}] \} \psi_k \, \mathrm{d}v \right\rangle ,$$

must then be false, since the wave functions ψ_k etc., do not cohere over even one wave length of the lattice wave. Quantum mechanically, we scarcely know what to do.

But now a macroscopic kinetic description is possible. Squeezing or distorting the lattice upsets the electron gas, which then buzzes around, is scattered by impurities etc., until it is brought into equilibrium.

The electron distribution will always lag a bit behind the lattice distortion, so that energy will be fed into it and degraded into heat. That is, the lattice wave is absorbed. If it is an ultrasonic wave, this absorption can be directly measured.

The interesting thing is to try to come back from this description towards shorter phonon wavelengths. This has been done in an ingenious paper by PIPPARD [16]. Using the same kinetic arguments, he has derived a formula for the m.f.p. of a phonon in a metal, which gives exactly the quantum mechanical answer in the limit $q\Lambda_e \gg 1$. It turns out that this formula is a function

chiefly of $q\Lambda_e$. When $q\Lambda_e > 1$ the quantum formula holds. When $q\Lambda_e < 1$ the macroscopic formula is correct.

The calculation of the resistance to lattice conduction because of the scattering of phonons by electrons must be adjusted for this phenomenon. The correction is possibly observable.

This seems to answer an old difficult question; when can we use perturbation theory and the adiabatic approximation? According to PEIERLS [2] these might break down unless $v_e\tau \geqslant 1$, i.e. when an electron state does not live as long as one period of the lattice. A much less stringent criterion was proposed by LANDAU, which amounts in effect to $k\Lambda_e > 1$ and this is nearly always satisfied. Now we have seen an actual calculation where a change of formula occurs when $q\Lambda_e \sim 1$. As suggested, this is the proper criterion for the adiabatic approximation. The point is that an electron does not stand still. It actually moves through the lattice with high velocity. Thus, it can adapt itself adiabatically to conditions in the wave, as it moves, and does not have to wait for the wave to pass it by.

Except for those few electrons which can « surf-ride », all different phases of the lattice wave, will be sampled many times in a m.p.f. and local conditions adjusted accordingly. The insistence that an electron should travel far enough ro see at least a whole wave is thus much more plausible than that it must last long enough for a whole wave to pass it by.

REFERENCES

[1] R. PEIERLS: *Ann. der Phys.* (5), **3**, 1055 (1929).
[2] R. PEIERLS: *Quantum Theory of Solids* (Oxford, 1955).
[3] J. M. ZIMAN: *Canad. Journ. Phys.*, **34**, 1256 (1956).
[4] P. G. KLEMENS: *Proc. Roy. Soc.*, A **208**, 108 (1951).
[5] P. G. KLEMENS: *Handb. der Phys.*, **14**, 198 (1956).
[6] P. G. KLEMENS: *Proc. Phys. Soc.*, A **68**, 1113 (1955).
[7] F. R. N. NABARRO: *Proc. Roy. Soc.*, A **209**, 278 (1951).
[8] H. G. B. CASIMIR: *Physica*, **5**, 595 (1938).
[9] H. STEHLE and A. SEEGER: *Zeits. f. Phys.*, **146**, 217, 242 (1956).
[10] S. C. HUNTER and F. R. N. NABARRO: *Proc. Roy. Soc.*, A **220**, 542 (1953).
[11] E. H. SONDHEIMER: *Adv. in Phys.*, **1**, 1 (1952).
[12] J. BARDEEN: *Phys. Rev.*, **52**, 688 (1937).
[13] F. BLOCH: *Zeits. f. Phys.*, **52**, 555 (1928).
[14] A. H. WILSON: *The Theory of Metals* (Cambridge, 1953), 2nd Edition.
[15] J. M. ZIMAN: *Proc. Roy. Soc.*, A **226**, 436 (1954).
[16] A. B. PIPPARD: *Phil. Mag.* (7), **46**, 1104 (1955).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Theory of Superconductivity.

J. R. SCHRIEFFER
University of Illinois - Urbana, Ill.

I would like to describe some recent work on the theory of superconductivity which Dr. J. BARDEEN, Dr. L. N. COOPER and I have carried on at the University of Illinois [1].

The main facts which a theory of superconductivity must explain are:

1) a second order phase transition at the critical temperature T_c;

2) an electronic specific heat varying as $\exp[-T_0/T]$ near $T = 0$ °K and other evidence for an energy gap for individual particle-like excitations;

3) the dependence of T_c and H_c, the critical magnetic field for destroying superconductivity, on isotopic mass: $H_c \sim T_c \sim 1/\sqrt{M_{\text{ion}}}$;

4) the Meissner-Ochsenfeld effect ($B = 0$) and effects associated with infinite conductivity ($E = 0$).

There are several laws of similarity which most superconductors obey, e.g. $\gamma T_c^2/H_0^2 = \text{const.}$, where γT is the electronic specific heat in the normal state and H_0 is the critical field at $T = 0$ °K. Such laws indicate that a universal theory should give an accurate description of these materials and only the gross features of the vibrational and electronic structure are of importance in accounting for the above facts.

The Sommerfeld-Bloch individual particle model, in which each electron is assumed to move independently in a self consistent field due to the other electrons and the ions gives a fairly good description of the normal state but fails to account for superconductivity. There is a large energy associated with correlation effects, however most of this energy occurs in both the normal and superconducting phases and cancels in the energy difference.

A major problem in constructing a satisfactory theory has been to isolate that part of the interaction which is responsible for the transition. The discovery of the isotope effect indicates, as had been independently suggested by FRÖHLICH [2], that the phonon electron interactions are primarily responsible for superconductivity. Early theories of Fröhlich [2] and Bardeen [3] based on the self energy of the electrons in the phonon field give the isotope effect correctly but do not yield a phase with superconducting properties and furthermore, the calculated condensation energy is far too large.

The superconducting phase presumably comes about by a resolution of the near degeneracy associated with small electronic excitations from the Fermi surface. This resolution is induced by some portion of the phonon-electron interaction and the screened Coulomb repulsion. The states which are pushed down by the interaction are qualitatively different from those observed in the normal phase in that a large number of electrons act coherently to give rise to the remarkable properties associated with the superconducting phase. It is this coherency which forms an energy gap for single particle-like excitations from the ground state. Such coherent states can not be treated by perturbation theory starting from the Bloch scheme and thus it is necessary to isolate that portion of the Hamiltonian which resolves the near degeneracy and treat this « reduced » Hamiltonian in an exact manner. The rest of the Hamiltonian will contribute essentially the same energy to both the « coherent » and the « normal » states and thus will not contribute to the condensation energy, nor, as it turns out, to other properties of interest in the superconducting phase.

Our theory is based on a collective ion-electron treatment given by BARDEEN and PINES [4], in which a true interaction between electrons due to virtual exchange of phonons, H_2, is derived in addition to the usual screened Coulomb repulsion, H_{Coul}. In their treatment, the diagonal terms of H_2 renormalize the single electron energies while the other terms in the Hamiltonian renormalize the phonon energies and describe the plasma modes. The interaction H_2, which we will call the phonon interaction, has matrix elements of the form

$$\langle k_2', k_1' | H_2 | k_2, k_1 \rangle = \frac{2 |M_\sigma|^2 \hbar\omega}{(\varepsilon_{k_1} - \varepsilon_{k_1'})^2 - (\hbar\omega)^2}, \qquad \sigma = k_1 - k_1'. \tag{1}$$

where M_σ is the matrix element for phonon-electron interaction calculated for zero-point amplitude of lattice vibrations, $\hbar\omega$ is the average phonon energy $\sim k\Theta_D$ and ε_k is the Block energy measured relative to the Fermi energy. For $|\varepsilon_k - \varepsilon_{k'}| < \hbar\omega$, H_2 gives an attractive interaction and it is this attraction which we believe to give rise to the condensed or superconducting phase (*).

(*) *Note added in proof.* – A new expression for the phonon interaction, which is free from singularities, has been derived by a variational method on the Hamiltonian including phonon variables. The interaction gives support to cutting off H_2 for $|\varepsilon_k| > \hbar\omega$ and is in agreement with the work of N. N. BOGOLJUBOV (to be published).

In our theory, only those terms in the Hamiltonian are kept which we believe to be responsible for the transition. The single particle energies and phonon frequencies are assumed to be temperature independent as are effects coming from transitions outside the region $|\varepsilon_k| < \hbar\omega$. While these assumptions have not been completely justified at the present time, the success of the theory indicates they are valid, although such effects could be included in the present formalism. The ground state is given by a superposition of many occupation number configurations (Slater determinants) which have the property that single electron states of opposite momentum and spin are occupied in pairs. It is this pairing that ensures that matrix elements of the interaction are of a fixed sign for those transitions of importance in the ground state, $|\varepsilon_k - \varepsilon_{k'}| \sim kT_c \ll \hbar\omega$ and allows for coherent mixing of configurations. The theory gives a criterion for superconductivity to exist which may be expressed as

$$-V = \left\langle -\frac{2|M_\sigma|^2}{\hbar\omega} + H_{\text{Coul}} \right\rangle_{\text{av.}} < 0\,, \tag{2}$$

where the average is taken over transitions of importance mentioned above. PINES [5] has been able to essentially reproduce Matthias' rules from (2) using a simple model with Thomas-Fermi screening. Physically, the superconducting phase will occur whenever the attractive phonon interaction dominates the screened Coulomb repulsion for those transitions for which $|\varepsilon_k - \varepsilon_{k'}| \leqslant \hbar\omega$.

In order to motivate our choice for the ground state function we observe that if a Hamiltonian has off-diagonal elements of predominantly negative sign, then a low energy state can be formed by a linear combination of the basis functions with expansion coefficients of predominantly one sign. In the actual problem, however, matrix elements of the interaction Hamiltonian generally alternate in sign due to the Fermi-Dirac statistics and a low coherent state is most readily formed by restricting the set of configurations to those giving matrix elements of one sign. Such a set is formed by those configurations in which states are occupied in pairs, that is if a state labeled by wave vector k_1 and spin σ_1 is occupied, so is its mate k_2, σ_2. For the ground state the best choice is $(k\uparrow, -k\downarrow)$ since this pairing gives the maximum amount of phase space for transitions and the antiparallel spin state gives a larger interaction energy than the parallel spin state due to absence of exchange. A state with net current flow can be constructed from pairs $(k\uparrow, -k+q\downarrow)$.

We start then by considering a problem in which states are occupied in pairs $(k\uparrow, -k\downarrow)$. It is convenient to work in a second quantized formalism with creation and annihilation operators for electrons which are based on the renormalized Bloch states and satisfy the usual Fermi anticommutation relations:

$$[c_{k\sigma}, c^*_{k'\sigma'}]_+ = \delta_{kk'}\delta_{\sigma\sigma'}\,, \tag{3}$$

$$[c_{k\sigma}, c_{k'\sigma'}]_+ = 0\,. \tag{4}$$

We define creation and annihilation operators for electron pairs as

$$b_k^* = c_{k\uparrow}^* c_{-k\downarrow}^*, \qquad b_k = c_{-k\downarrow} c_{k\uparrow}, \tag{5}$$

with the wave vector of the spin up member labeling the pair.

The « reduced » Hamiltonian which is essentially responsible for the superconducting state is

$$H_{\text{red}} = \sum_{(k>k_F,\,\sigma)} \varepsilon_k n_{k\sigma} + \sum_{(k<k_F,\,\sigma)} |\varepsilon_k|(1 - n_{k\sigma}) - \sum_{(kk')} V_{kk'} b_{k'}^* b_k \,, \tag{6}$$

where the first two terms give the Bloch energy of the electrons and holes and the interaction is defined by

$$-V_{kk'} = (-k'\downarrow, k'\uparrow | H_2 + H_{\text{Coul}} | -k\downarrow, k\uparrow) + (k'\uparrow, -k'\downarrow | H_2 + H_{\text{Coul}} | k\uparrow, -k\downarrow) . \tag{7}$$

We use a variational approach to determine the ground state energy of H_{red} with the trial function chosen to be

$$\Psi_0 = \prod_k (\sqrt{1 - h_k} + \sqrt{h_k} b_k^*) \Phi_0 \,, \tag{8}$$

where Φ_0 is the vacuum, b_k^* creates a pair in state k and h_k is the distribution function for these « virtual » pairs. The function Ψ_0 does not describe a system of fixed number of particles, however the fluctuations are extremely small and lead to no difficulty. It is easily seen that the probability for a configuration having pair states $k_1, \ldots, k_n$ occupied and $k_1', \ldots, k_n'$ empty in Ψ_0 is simply $\prod_k h_k \prod_{k'} (1 - h_{k'})$ which resembles the intermediate coupling approximation.

The ground state energy relative to the energy of the renormalized Fermi sea is given by combining (8) and (6) and using the commulation relations (3) and (4).

$$W_0 = (\Psi_0, H_{\text{red}} \Psi_0) = 2 \sum_{k>k_F} \varepsilon_k h_k + 2 \sum_{k<k_F} |\varepsilon_k|(1 - h_k) - \\ - \sum_{k,k'} V_{kk'} \sqrt{h_k (1 - h_k) h_{k'} (1 - h_{k'})} \,. \tag{9}$$

The first two terms give the Bloch energy of the electrons and holes with the factor of two accounting for two electrons in the pair. It is instructive to derive the last term from a probabilistic interpretation of Ψ_0.

Fig. 1 shows a typical transition in which a pair in k goes to k'. This transition is possible only if k is occupied and k' unoccupied in the initial state and the reverse is true for the final state, all other pairs remaining fixed. The probability for this to be true is $[h_k(1-h_k)\,h_{k'}(1-h_{k'})]$. Since it is probability amplitudes rather than probabilities which enter in matrix elements, the interaction energy is given by summing the matrix element weighted by the square root of this factor over all possible transitions.

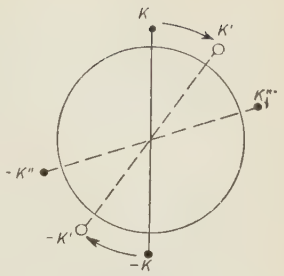


Fig. 1.

The distribution function for virtual pairs is determined by minimizing W_0 with respect to h_k and one finds

$$\frac{\delta W_0}{\delta h_k} = 0\,;\quad \frac{\sqrt{h_k(1-h_k)}}{1-2h_k} = \frac{\sum_k V_{kk'}\sqrt{h_{k'}(1-h_{k'})}}{2\varepsilon_k}\,. \tag{10}$$

The integral equation is easily solved if $V_{kk'}$ is replaced by a constant average matrix element V for those transitions with $|\varepsilon_k|$ and $|\varepsilon_{k'}| < \hbar\omega$ and by zero outside this region. The average is primarily one over directions in k-space since $V_{kk'}$ is relatively insensitive to energy in the region of importance. If the average matrix element approximation is used, we find:

$$h_k = \begin{cases} \dfrac{1}{2}\left(1-\dfrac{\varepsilon_k}{E_k}\right) & |\varepsilon_k| < \hbar\omega\,, \\ 0 & |\varepsilon_k| > \hbar\omega\,, \end{cases} \tag{11}$$

where

$$E_k = +\sqrt{\varepsilon_k^2+\varepsilon_0^2} \tag{12}$$

and

$$\varepsilon_0 = V\sum_k \sqrt{h_k(1-h_k)} = \frac{\hbar\omega}{\sinh\left[1/N(0)V\right]}\,, \tag{13}$$

N(0) being the density of renormalized Bloch states of one spin at the Fermi surface. As we shall see the energy gap for single particle excitations is $2\varepsilon_0$ and E_k is the energy for electrons in what corresponds to the « normal » component of the two fluid model. If we combine (8), (11), (12) and (13), we find the condensation energy is given by

$$W_0 = \frac{-2N(0)(\hbar\omega)^2}{\exp\left[2/N(0)V\right]-1} = -\frac{H_0^2}{8\pi}\,, \tag{14}$$

which agrees with the isotopic effect since $N(0)$ and V are independent of isotopic mass and $H_0^2 \sim (\hbar\omega)^2 \sim 1/M_{\text{ion}}$. The criterion (2) follows because the condensed state does not exist for $V \geqslant 0$. Empirically, $W \sim -2N(0)(kT_c)^2$ and therefore

$$\left(\frac{kT_c}{\hbar\omega}\right)^2 \sim \frac{1}{\exp[2/N(0)V] - 1}\,, \tag{15}$$

however $kT_c \ll \hbar\omega$ in general and thus $N(0)V < 1$, or the weak coupling case holds. Therefore, T_c is much smaller than the Debye temperature because of the cancellation of the phonon and Coulomb interactions between the electrons. It should be noted that W_0 could not have been obtained from a perturbation expansion.

The effects of neglected terms in the Hamiltonian, such as interactions which lead away from $(k\uparrow, -k\downarrow)$ configurations, have been estimated and it appears that their contribution to W_0 is quite small.

Excited states.

The simplest excited state is formed by breaking up a pair in state k', the spin up member going to $k'\uparrow$. By making an argument identical to that used in calculating the interaction energy for W_0, we see that the number of transitions is reduced by about $2N(0)\varepsilon_0$ since transitions can be made to neither k' nor k'' by pairs in other states. Since each transition contributes an energy $\sim V$, we find the energy required to form a single particle-like excitation is $\sim 2\varepsilon_0 N(0)V \sim 2\varepsilon_0$, which we shall see is of the order of $3.5\,kT_c$. It is the high coherency of the ground state wave function which results in the condensation energy being proportional to the square of the number of virtual pairs and gives the energy gap. Detailed calculation gives the excitation energy as $E_{k'} + E_{k''}$ for this excitation.

The complete set of excited states is given by specifying those pair states which are occupied by single particles (either $k\uparrow$ or $-k\downarrow$ occupied, but not both) and those occupied by real pairs. By a real pair function we mean a function which is orthogonal to the virtual pair function for that state. Real pairs are represented by the operator

$$(\sqrt{1-h_k}\,b_k^* - \sqrt{h_k}) \tag{16}$$

as opposed to the virtual pair operator

$$(\sqrt{1-h_k} + \sqrt{h_k}\,b_k^*)\,. \tag{17}$$

A typical excited state can be represented by

$$\Psi_{\text{exc}} = \prod_{k(J)} (\sqrt{1-h_k} + \sqrt{h_k}\, b_k^*) \prod_{k'(P)} (\sqrt{1-h_{k'}}\, b_{k'}^* - \sqrt{h_{k'}}) \prod_{k''(S)} c_{(k'')}^* \Phi_0 \,, \tag{18}$$

where (k') means either $k''\uparrow$ or $-k''\downarrow$, is included in the product and S and P specify the states occupied by singles and reals with the rest of the states in the region $|\varepsilon_k| < \hbar\omega$, J being occupied by virtuals. To determine the distribution of the excited particles at any temperature, the free energy is minimized with respect to the distribution as well as to h_k. After some calculation one finds that the distribution of single particles S_k and real pairs P_k is given by

$$S_k = 2f_k(1-f_k)\,, \tag{19}$$

$$P_k = f_k^2\,, \tag{20}$$

where

$$f_k = \frac{1}{\exp[E_k/kT]+1}\,. \tag{21}$$

The energy gap is determined by

$$\frac{1}{N(0)V} = \int_0^{\hbar\omega} \frac{\mathrm{d}\varepsilon}{\sqrt{\varepsilon^2+\varepsilon_0^2}} \operatorname{tgh}\left(\frac{\sqrt{\varepsilon^2+\varepsilon_0^2}}{2kT}\right) \tag{22}$$

and is plotted as a function of temperature in Fig. 2. From the form of (19), (20) and (21), it follows that the single particles and real pairs together describe

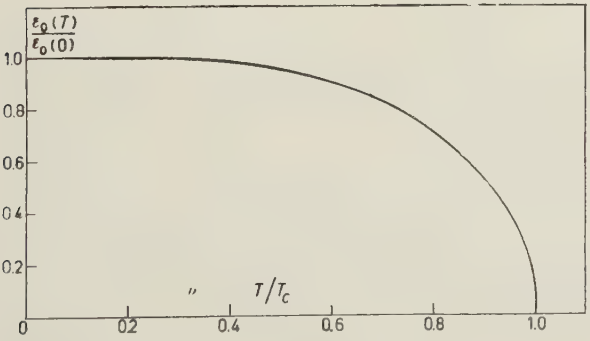


Fig. 2.

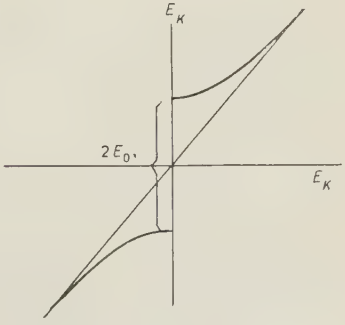


Fig. 3.

a set of non-interacting fermions with the dispersion law (12). These excited electrons (and the corresponding holes below the Fermi surface) are to be identified with the normal component of the two-fluid model. The modified

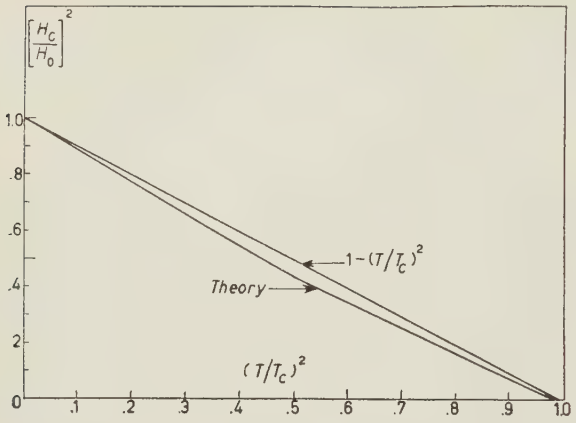


Fig. 4.

energy spectrum is shown in Fig. 3 where an energy gap of width $2\varepsilon_0$ appears centered about the Fermi surface. At the transition temperature the gap vanishes and from (22) we have

$$(23)\quad \begin{cases} kT_c = \hbar\omega \exp\left[-\dfrac{1}{N(0)V}\right] \\ \text{or} \\ \dfrac{2\varepsilon_0}{kT_c} = 3.50\,, \end{cases}$$

which is in agreement with the isotope effect.

Above the transition temperature, our theory goes over to the usual single particle Bloch scheme for the normal state. Thus our treatment is an extention of the Bloch scheme to the superconducting phase since the single particles are treated in the single particle approximation at all temperatures.

The critical field curve predicted by our theory is shown in Fig. 4 and shows the observed deviation from the two-fluid model ($I - t^2$) law. The electronic specific heat at low temperatures is of the form

$$(24)\quad \frac{c_{es}}{\gamma T_c} = 8.5 \exp\left[-\frac{1.44\,T_c}{T}\right],$$

and the theory explicitly gives a second order phase transition with a jump in specific heat given by

$$(25)\quad \left.\frac{c_{es}}{\gamma T_c}\right|_{T_c} = 0.170\,,$$

both of which are in reasonable agreement with experiment.

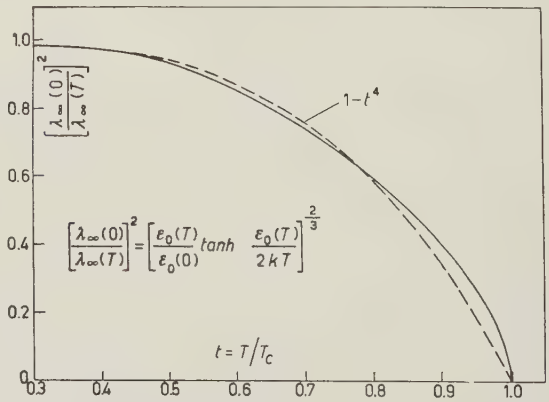


Fig. 5.

The non-local relation between the current density and the electromagnetic vector potential is essentially of the form suggested by PIPPARD [6] on empirical grounds and the penetration depth is plotted as a function of temperature in Fig. 5.

Boundary effects and several transport properties are still be to treated at present and it would be desirable to have an improved general formulation of the theory (*).

(*) A set of variables which is convenient for generating the excited states described above has been introduced by J. VALITIN (to be published) and N. N. BOGOLJUBOV (to be published). It has been possible to obtain an improved general formulation of the theory, including the variation of the free energy, the energy gap, etc. with temperature, in terms of these variables. Details of this work will be published elsewhere.

REFERENCES

[1] J. BARDEEN, L. N. COOPER and J. R. SCHRIEFFER: *Phys. Rev.*, **106**, 162 (1957) and to be published.
[2] H. FRÖHLICH: *Phys. Rev.*, **79**, 845 (1950).
[3] J. BARDEEN: *Phys. Rev.*, **79**, 167 (1950). For a review of early work see J. BARDEEN: *Rev. Mod. Phys.*, **23**, 261 (1951).
[4] J. BARDEEN and D. PINES: *Phys. Rev.*, **99**, 1140 (1950).
[5] D. PINES: to be published.
[6] A. B. PIPPARD: *Proc. Roy. Soc. London*, A **261**, 547 (1953).

PARTE SECONDA

Difetti reticolari nei solidi.

Dislocation Theory.

F. C. Frank

H. H. Wills Physical Laboratory - University of Bristol

1. – Introduction.

The main aim of these lectures is to introduce most of the main concepts of dislocation theory, picking on particular examples to illustrate them, with no attempt at completeness.

Before direct observations could be made on dislocations there were four main classes of experimental evidence pointing to the existence of a class of crystal imperfections, which we now analyse in terms of the dislocation concept. The earliest body of evidence was that from X-ray diffraction: within a year or two of the discovery of this phenomenon by Laue, Friedrich and Knipping in 1912, it became evident that the diffraction intensities from most real crystals were very different from those one would calculate on the assumption of ideal crystal structures. A rather ill-defined concept of « mosaic structure » was employed for a generation to account in general terms for these discrepancies. It is only within the last few years that we have achieved a well-defined picture, for some cases, of what this mosaic structure really is: we are able to describe it now in some detail, in terms of dislocations. The next body of evidence was concerned with the mechanical properties of solids, which, to put it briefly, are a thousand times weaker than ideal crystals. The fact became increasingly evident as people acquired confidence in the Born type of model for the ideal crystal, which accounts very well for the elastic or thermal properties, but indicates plastic strengths vastly in excess of those of technical materials. It was to explain these discrepancies that Taylor and Orowan independently put forward the concept of crystal dislocations in 1934. Related ideas had been employed by Prandtl, Dehlinger, Polanyi, and Smekal at earlier dates, but the modern development starts very clearly from these two papers in 1934. It was followed by 15 years or so of theoretical development, with no *direct* experimental observation of dislocations, leading to a cleavage between metal physicists who employed the dislocation concept and many metallurgists,

who regarded it as a complex fiction introduced to rescue an unsuccessful theory. A third body of evidence was latent in the literature—a very large discrepancy between observed rates of crystal growth and rates calculated on the assumption that crystals were ideal. The very large quantitative discrepancy was obscured by a qualitative similarity between observed and calculated laws of dependence of growth rate on supersaturation. Once recognized it became another experimental fact attributable to the presence of dislocations in real crystals and very soon led to the clearest visual evidence that they were present. By this time, the study of point defects had also given compelling evidence that the crystal surface was not the only place at which the number of crystal lattice sites could alter—for example, the rate of formation of *F*-centres in alkali halides under irradiation was too great to suppose that vacancies could only form at the surface. This field, the interrelationship of dislocations with point defects, is of course the field in closest relationship with other contributions to this Summer School; it is also the field which has provided in very recent years the most detailed observational evidence that we have about the actual arrangements of dislocations in real crystals.

Dislocation theory falls into parts of various status. There are some conclusions which are exact, because they are purely geometrical or topological. There are some parts which though approximate are nevertheless very reliable (for example, the conclusion that dislocations should be absent from a crystal in thermodynamic equilibrium — thermodynamic properties of dislocations have only been roughly estimated, but nevertheless the error could not be large enough to reverse the conclusion). And then, naturally, many of the questions of current interest are those for which one cannot be sure of the reliability of approximations made, and must draw on all the resources of theory, observation, and guesswork in the attempt to make further progress. Above all, as I see it, the theory of the work hardening of metals, the subject in which the intensive study of crystal dislocations began, remains very definitely in this uncertain stage.

Fig. 1. – Edge dislocation.

The most familiar picture of a dislocation is Taylor's picture of the edge dislocation (Fig. 1). This picture has to be completed three-dimensionally, the dislocation line being perpendicular to the plane of the drawing, at the edge of the incomplete lattice plane. Though easiest to draw, the edge dislocation is not the simplest one, which is rather the screw dislocation, the significance of which was first pointed out by BURGERS in 1939. To describe the state of strain at a screw dislocation, we picture (Fig. 2) a hollow cylinder, cut along a radial surface, at which the two cut faces are given an axial relative displacement of amount b. Then

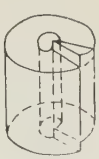


Fig. 2. – Screw dislocation.

every element of a cylindrical shell of thickness $\mathrm{d}r$ at radius r in this cylinder has a shear strain

$$\gamma = \frac{b}{2\pi r} .$$

The cylindrical surfaces of such a shell are free from stress, so that each shell is in equilibrium within itself, provided the stresses are matched by suitable stresses at the end surfaces: we assume for the moment that these are present. The shear involves an energy density

$$\tfrac{1}{2}\mu\gamma^2 = \tfrac{1}{2}\mu b^2/4\pi^2 r^2 ,$$

where μ is the shear modulus. Hence the elastic energy per unit length in a cylinder of internal radius r_0 and external radius R is

$$E = \int_{r_0}^{R} \left(\frac{1}{2}\mu b^2/4\pi^2 r^2\right) 2\pi r \,\mathrm{d}r = \frac{\mu b^2}{4\pi} \ln \frac{R}{r_0} .$$

In application to crystals, we take b to be an interatomic spacing and the inner radius r_0 to be a distance of the same order. This condition is not perfectly in equilibrium, without applied tractions at the ends. These have the magnitude

$$\tau = \mu\gamma = \mu b/2\pi r ,$$

producing a torque which is found by integration to be

$$\tfrac{1}{2}\mu b(R^2 - r_0^2) \sim \tfrac{1}{2}\mu b R^2 .$$

If these tractions are absent, there will be additional displacements produced by negative tractions of the same magnitude. Except within a few diameters of the ends these displacements are to be described as a simple torsion produced by the given torque. This amounts to

$$\mathrm{d}\theta/\mathrm{d}z = b/\pi R^2 .$$

This torsion is not of much importance when R is large, but in the case of a thin metal « whisker » which may have a screw dislocation along its axis it becomes a large, easily observed quantity (e.g., about 40° per cm if $b = 2.5 \cdot 10^{-8}$ cm and $R = 10^{-4}$ cm). It has been shown by Eshelby that (rather surprisingly) the screw dislocation will be stable in this position. This

torsion has been looked for in the Bell Laboratories by X-ray diffraction, but thus far only in tin whiskers for which the mode of formation would not lead me to expect an axial screw dislocation to be present.

2. – One-dimensional models.

The first purpose of introducing the idea of crystal dislocations was to explain plastic deformation. Simple consideration of Fig. 1 indicates that the dislocation ought to be relatively easily displaced through the crystal, producing a plastic deformation. The model of continuum elasticity with which we have discussed the screw dislocation is of little value, directly, for discussing dislocation mobility.

We study now some one-dimensional models having properties related to those of dislocations.

1) A heavy rope lying on corrugated iron.

2) A sinusoidal surface on which a series of weights is put, the weights being connected by springs (a model considered by FRENKEL and KONTOROVA).

If α is either the tension in the rope or the modulus of the springs then we have for the displacement u from an equilibrium position (the substrate period in u being 2π) the following equations

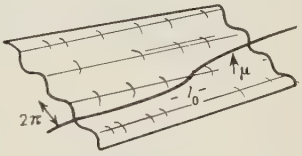


Fig. 3.

$$(1) \qquad \alpha \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} = \beta \sin u \quad (x, \text{co-ordinate along the rope}),$$

$$(2) \qquad \alpha(u_{m+1} - u_m) - \alpha(u_m - u_{m-1}) = \beta \sin u \; ;$$

the second difference may be approximated by $\mathrm{d}^2u/\mathrm{d}x^2$, in which case both equations are identical. These equations are of the form

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2} = y \sin u \; . \qquad y = \beta/\alpha.$$

The general solution is expressed by elliptic integrals (representing sequences of « dislocations ») but we may be content with the solution representing a single dislocation:

$$\operatorname{tg} \frac{u}{4} = \exp\left[-y^{\frac{1}{2}} x\right] .$$

The effective length of the dislocation may be defined as

$$l_0 = \frac{2\pi}{(\mathrm{d}u/\mathrm{d}x)_{\max}} = \pi \Big/ \frac{\alpha}{\beta} .$$

If we use this model to represent a monolayer of atoms on a crystal surface (Fig. 5), and choose the constants of the equations to agree with interatomic forces of the form $-ar^{-7}+br^{-13}$, we find that l_0 is about 7 interatomic

Fig. 4.

Fig. 5.

spacings. Either version of our model, corresponding to Fig. 4 or 5 (or alternative equivalent models) readily exhibit the mobility of dislocations. To include the possibility of moving, the foregoing equation is to be replaced by

$$\alpha\frac{\partial^2 u}{\partial x^2}-\varrho\frac{\partial^2 u}{\partial t^2}=\beta\sin u\,,$$

which becomes

$$\frac{\partial^2 u}{\partial x^2}-\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}=y\sin u\,,$$

with $c=(\alpha/\varrho)^{\frac{1}{2}}$ the velocity of sound in the rope, or chain of springs and masses, in the absence of the periodic substrate. For the case of steady motion we may solve this equation by introducing a new variable

$$X''=\frac{x-vt}{\sqrt{1-v^2/c^2}}\,.$$

This reduces the equation to the same form as we had for the static case, the effective length being reduced by a « Lorentz contraction » factor $\sqrt{1-\beta^2}$, with $\beta=v/c$. One can also calculate the energy

$$E=\frac{1}{2}E_0\left\{\frac{1-\beta^2}{\sqrt{1-\beta^2}}+\frac{1}{\sqrt{1-\beta^2}}+\frac{\beta^2}{\sqrt{1-\beta^2}}\right\}=\frac{E_0}{\sqrt{1-\beta^2}}\qquad\left(\beta=\frac{v}{c}\right).$$

consisting of three parts, the first two being the potential energy of the weights or the rope brought out of their equilibrium position at the bottom of groove or well and the work done against tension in rope or springs, and the third part the proper kinetic energy. We have a contracted strain field and an increased energy when the dislocation moves.

It can be shown that the same formulae apply for three-dimensional dislocations—accurately for screw dislocations, in an isotropic medium, and qualitatively in general, when we have the complication that more than one speed of sound is involved. As the dislocation approaches the speed of sound

its energy approaches infinity. It is believed that in ordinary slip processes dislocations do not move faster than $c/10$, but they may move considerably faster in some processes of twinning or martensitic transformation, so that a fully developed theory of these processes may need to take this « relativistic » increase of energy into account. For low velocity processes we may usefully derive the result that the effective mass of a dislocation line is equal to its rest energy divided by c^2.

3. – Peierls-Nabarro model.

In the first lecture, we used the linear theory of elasticity to describe the state of strain of a hollow cylinder. The shear strain $\gamma = b/2\pi r$ will amount to a few % if the radial distance r is about two or three times b. Crystals will take linearly strains of the order of 2 or 3 %. To take non-linearity into account, one can adopt the model proposed by PEIERLS and further developed by NABARRO, LEIBFRIED and DIETZE, ESHELBY, and VAN DER MERWE. We cut the crystal into two pieces, in which we assume linear elasticity theory to be valid, and assume a special type of connection between the two surfaces. If u_1 and u_2 are the displacements of corresponding points on the two surfaces, the tangential forces are of the form $A \sin(u_2 - u_1)$, the constant A being chosen so that the maximum slope of the graph of the restoring force corresponds to the rigidity modulus of the material acting in the gap of width a.

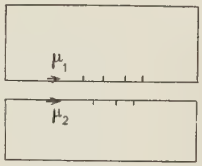


Fig. 6. – The Peierls-Nabarro model.

LEIBFRIED showed that if the gap a between the two surfaces is equal to the distance b between atoms along the slip direction, the strain-stress systems are identical with the strain-stress systems given by the continuous model. The same type of model may be used both for edge dislocations and for screw dislocations, differing only in the direction of the displacements relative to the dislocation line. The energy per unit length of the dislocation is

$$W_0 = \mu \frac{b^2}{4\pi K} \ln \frac{Kr}{a},$$

where μ is the rigidity modulus, $K = 1$ for screw dislocations and $K = 1 - \nu$ for edge dislocations (ν, Poisson's ratio). Thus K is between 1 and $\frac{1}{2}$, or between 1 and 0.7 for most materials.

An important question for the mobility of dislocations is how much the energy fluctuates as the dislocation moves. During its motion, an edge dislocation passes from the symmetrical configuration 1 to the different symmetrical configuration 2. The Peierls model gives the result that the energy

in the first configuration and the energy in the second configuration are identical, and that as the dislocation moves the energy fluctuation has a maximum between the two symmetrical configurations. This is a consequence of the simple sinusoidal form assumed for the non-linear interaction. HUNTINGTON has discussed this point, and proposed some refinements. Personally, I have no great faith in any *quantitative* calculation of the Peierls stress, the stress needed to make the dislocation move, but consider that the value of the calculation is to exhibit in principle the great reduction in yield stress due to the presence of dislocation. According to the simple theory the amplitude of the energy fluctuation is

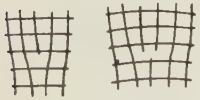


Fig. 7. – Two symmetrical positions of an edge dislocation.

$$\Delta W = \mu \frac{b^2}{2\pi K} \exp\left[-\frac{2\pi a}{Kb}\right],$$

and the corresponding critical shear stress is

$$\sigma = \frac{2\mu}{K} \exp\left[-\frac{2\pi a}{Kb}\right],$$

For some crystallographic planes in the f.c.c. lattice one has (taking a to be the interplanar spacing)

Plane	a/b	σ_c (*)
(111)	$\sqrt{2/3}$ or $\sqrt{2}$ (+)	$3 \cdot 10^{-3}$ μm or 10^{-5} μm
(100)	$1/\sqrt{2}$	$6 \cdot 10^{-3}$ μm
(110)	$1/2$	$3 \cdot 10^{-2}$ μm

These are not unreasonable figures.

For well annealed single crystals of f.c.c. metals, there is some degree of plastic yield at $\sigma \sim 10^{-5}$ μm. The slip is on the (111) plane. At higher stresses, slip occurs also in the cube planes. Without dislocations critical stresses of the order 10^{-3} μm would be expected.

4. – Edge dislocation in a cylinder.

We assume that the displacements are in a plane. We may then make use of Airy functions. Any function $\chi(x, y)$ satisfying the equations $\nabla^4 \chi = 0$ represents a state of equilibrium in plane strain, for isotropic elasticity, ac-

(*) J. FRIEDEL reports, in his book *Les dislocations*, values which are a thousand times lower than those reported here.

(+) The value $\sqrt{2}$ refers to the splitting of dislocations, to be defined later.

cording to the prescription:

$$\sigma_{xx} = \frac{\partial^2 \chi}{\partial y^2}, \qquad \sigma_{yy} = \frac{\partial^2 \chi}{\partial x^2}, \qquad \sigma_{xy} = -\frac{\partial^2 \chi}{\partial x \partial y}.$$

With the assumption $\chi = R(r) \cdot \Theta(\theta)$, and adopting as a trial solution $\Theta(\theta) = \sin\theta$, one has

$$\chi = A\left(r \ln r + \frac{r_0^2}{2r} - \frac{r^3}{2R^2}\right) \sin\theta,$$

where A, r_0 and R are constants chosen to give zero stresses on these inner and outer cylindrical surfaces. The essential term describing the dislocation is the logarithmic term. As a result, one has a compression in the half crystal which contains the extra half plane of the dislocation and a dilatation in the other half crystal, the two half crystals being separated by the slip plane of the dislocation. Calculations made with anisotropic elasticity have not so far shown that any practically important conclusion is qualitatively altered. There is probably not much point in introducing this complication without at the same time taking account of the discreteness of the crystal and the details of structure at the core of the dislocation.

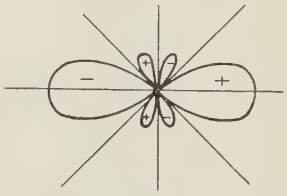


Fig. 8.

Consideration of the stresses at the ends shows that an edge dislocation along the axis should not produce any bending of the cylinder, analogous to the torsion which occurs with screw dislocations.

The resulting stress field (omitting the surface terms) is most simply described in polar co-ordinates:

$$\sigma_{rr} = \sigma_{\theta\theta} = \frac{-\mu b}{2\pi(1-\nu)} \cdot \frac{\sin\theta}{r},$$

$$\sigma_{\theta r} = \frac{\mu b}{2\pi(1-\nu)} \cdot \frac{\cos\theta}{r},$$

$$\sigma_{zz} = \frac{-2\mu b}{2\pi(1-\nu)} \cdot \frac{\sin\theta}{r}.$$

However, the most important stress is the shear stress, σ_{xy}, on planes parallel to the glide plane, $\theta = 0$. This is expressed by

$$\sigma_{xy} = \frac{\mu b}{2\pi(1-\nu)} \cdot \frac{\cos\theta \cos 2\theta}{r},$$

illustrated in Fig. 8. Of especial importance is the existence of planes of zero stress at 45°, 90°, etc., with sectors of reversed stress, leading to a position of metastable equilibrium for a second similar dislocation in the direction $\theta = \pi/2$.

The energy per unit length of the edge dislocation (apart from surface terms) is

$$E_0 = \frac{\mu b^2}{4\pi(1-\nu)} \ln\left(\frac{R}{r_0}\right).$$

The inner hole of radius r_0 is, thus far, only a device to eliminate the singularity. The appropriate value of r_0 for the energy formula can be estimated in various ways. According to the elastic theory, the energy density varies as r^{-2}; but Hooke's law fails when $r < \sim 5b$ where the strain is about 3%. It fails in such a manner that the energy density increases less rapidly. We may call the energy density within this radius the core energy. The Peierls-Nabarro model gives us one value for it. BRAGG estimated what should be an upper limit by assuming obedience to Hooke's law till the energy density reaches the latent heat of melting, and a constant energy density in the region closer to the centre. For the particular case of NaCl HUNTINGTON has estimated the core energy by direct consideration of ionic interactions according to a Born model of the crystal. Some experimental estimates can be obtained from « observed » values of the surface-tension of small-angle intercrystalline boundaries, which, as we shall see, are really systems of dislocations. By these various methods, we find that the core energy can be absorbed into the elastic expression by putting r_0 equal to about 1 or 2 times b.

Let us now consider the magnitude of R. Our cylinder is a rather artificial example, but we may solve the elastic problem for a variety of other external geometries, always obtaining a dominant energy term of the same form, with R of the order of magnitude of a macroscopic dimension of the situation—for example, the nearest distance between the dislocation and a free surface, the half-distance between a parallel pair of dislocations of opposite sign, or the radius of a dislocation loop. In these cases the stress field nearer to the dislocation than R is closely similar to that of an isolated dislocation in an infinite body, falling to zero at distances large compared with R. As a rule, no great precision is required in estimating R. If we say R lies between 10^{-4} and 10^{-1} cm, and r_0 between 0.5 and 5 Å, the range of variation of $\ln(R/r_0)$ is about a factor 2. It can often be treated as a constant without serious error. Its magnitude approximately cancels the divisor $4\pi K$ (K lying between about 0.7 and 1 according as the dislocation is screw or edge). Then we have $E_0 \sim \mu b^2$: or the energy per atomic length along the dislocation line is approximately μb^3. For elementary dislocations in metals this is of the order 1 eV.

5. – More general dislocation lines.

Fig. 9 is a picture deriving from J. M. BURGERS (1939), which introduced much more generality and flexibility into our thinking about dislocations. One should visualize the black dots as lying in front of the plane of the picture, with essentially similar patterns repeated in further planes in front of this, and the open circles lying behind the plane of the picture, with essentially ismilar patterns repeated in further planes behind this. Then we have an edge dislocation line leading forwards, out of the plane of the picture, connecting with a screw dislocation line leading to the right, in the plane of the picture. This screw dislocation lies in the slip plane of the edge dislocation, and the same plane can serve as a slip plane for the screw. From this we can proceed to the representation of a dislocation line which is of mixed edge and screw character, and lies along an arbitrary path in the glide plane—first, as a stepped line of edge and screw portions; then, in the limit, as a line of arbitrary orientation or continuous curvature. The centre of the dislocation line can be of more continuous curvature than the discrete lattice structure of the crystal because the dislocation is several atoms wide at the core.

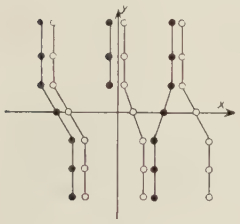


Fig. 9.

To the approximation of isotropic elastic theory, we obtain the stress and strain fields of a dislocation in arbitrary orientation by superposing those of edge and screw dislocation in appropriate proportions (given by $\sin\psi$ and $\cos\psi$, where ψ is the angle between slip direction and dislocation line). Such superposition is always permissible when Hooke's law is obeyed. In the present case, since the stresses of edge and screw dislocations are orthogonal, we can also add the energies (in proportions $\sin^2\psi$ and $\cos^2\psi$). Since, furthermore, the energies of edge and screw dislocations are not very different (roughly in the ratio 4 to 3) we can often disregard the variation of this orientation dependent factor as well as the variation of the factor $\ln(R/r_0)$ and proceed as though dislocation lines had a constant energy per unit length. This is the basis of the method introduced by MOTT and NABARRO for the approximate calculation of the elastic equilibrium of dislocation lines in complex configurations: in this method we regard them as dislocation lines having a constant line tension, equal to their energy per unit length.

An example of a general dislocation line is illustrated in Fig. 10. The dislocation line, which is edge

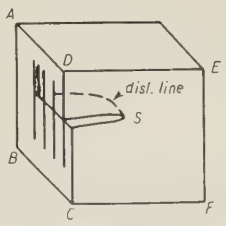


Fig. 10.

on the surface plane $ABCD$ and screw on the surface plane $CDEF$ transformed into a helicoid, is continuously changing in character in the intermediate region.

It is intuitively obvious that an applied shear stress will tend to expand the slipped area in all directions. We show later that the effective force per unit length on the dislocation in its glide plane is proportional to the magnitude of b and to the shear stress in the glide plane resolved in the slip direction. That this is true on the average is very simply shown by considering the work done when a dislocation moves right across a crystal.

6. – Flexible nature of dislocations.

Now let us consider a dislocation line, pinned at two points, e.g. by two impurity atoms. The dislocation line between the two points may oscillate like a stretched string. The line tension of the dislocation is

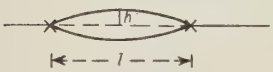


Fig. 11.

$$E_0 = \mu \frac{b^2}{4\pi K} \ln \frac{R}{r_0} ,$$

which in the Nabarro approximation is constant and $\sim \mu b^2$. The mass per unit length of the dislocation is E_0/c^2, and the speed of travel of a sound pulse along the dislocation line is

$$v \sqrt{\frac{E_0}{E_0/c^2}} = c ;$$

namely this speed is equal to the speed of sound in the crystal, to the approximation that the longitudinal and transverse velocities of sound are equal, as in the Debye approximation for the specific heats. If h is the amplitude of vibration, and l the half wavelength of the fundamental oscillation, one has

$$\frac{\pi^2 h^2}{l} E_0 = \frac{1}{2} kT ,$$

whence

$$\left(\frac{h}{b}\right)^2 = \left(\frac{kT}{2\pi^2 E_0 b}\right)\left(\frac{1}{b}\right) \sim 10^{-3} \left(\frac{1}{b}\right) .$$

Thus for a loop one thousand atoms long the oscillations are of about one atomic spacing.

7. – Entropy of a dislocation line.

As we have seen, the energy of a dislocation in typical metals is about 1 eV per atomic length. COTTRELL discussed the question whether a source of entropy can be found to compensate such a large energy. COTTRELL introduces the concept of the flexibility of a dislocation line: the dislocation can go from a fixed point in a lattice plane to z places in the next plane. If one assumes $z \sim 6$, and n is the number of the planes through which the dislocation passes one had a contribution to the free energy

$$\Delta F = T\Delta S \sim nkT \ln 6 \sim 2nkT .$$

As COTTRELL remarks this almost certainly overestimates the flexibility and should be regarded as an upper limit.

FRIEDEL considers alternatively the contribution to the free energy due to the oscillations of the dislocation lines. Since the speed of sound along the dislocation line is the same as in the crystal, the highest frequency in the spectrum of these oscillations is the maximum frequency of the Debye spectrum $\nu_m = k\Theta/h$.

FRIEDEL superposes this frequency and its harmonics $\nu_j = (j/n)\nu_m$ on the Debye spectrum. This must surely overestimate the contribution to the free energy of the crystal from this source since the normal modes of the crystal, dislocated or not, are $3N$ (N, number of atoms in the crystal). I would therefore argue that we must subtract n modes from the Debye spectrum. The contribution to the free energy per oscillator is:

$$F_{\nu_j} = \tfrac{1}{2}h\nu_j + kT \ln \left(1 - \exp\left[-h\nu_j/kT\right]\right)$$

and the total additional free energy due to the oscillations of the dislocation line, if we subtract the n modes from the top of the Debye spectrum (taking them from anywhere else has less effect) is

$$\Delta F = \sum_{j=1}^{n} \left\{ \frac{1}{2}\left(\frac{j}{n} - 1\right) k\Theta + kT \ln \frac{1 - \exp\left[-j\Theta/nT\right]}{1 - \exp\left[-\Theta/T\right]} \right\} .$$

For $T \to 0\,^{\circ}\mathrm{K}$ one has

$$\Delta F = -\frac{n-1}{4}\, k\Theta ,$$

that is about $-\frac{1}{4}k\Theta$ per atom along the dislocation. For $T \gg \Theta$ one has

$$\Delta F \sim -\tfrac{1}{4}nk\Theta - k\Theta .$$

As a result of these considerations one has that the free energy associated with a dislocation line is positive up to the melting point, except in the case of a very large density of dislocations. When dislocations are very close together (about 10 Å apart) the term $\log (R_1/r_0)$ becomes small and the free energy can be negative at sufficiently high temperatures. In this case the free energy may decrease as the number of dislocations is increased at high temperatures. This may provide a formal description of melting in terms of the dislocation density increasing spontaneously in a co-operative manner. It is not very useful however, because the whole concept of dislocations begins to lose its meaning when the dislocation density becomes very high, as we shall see when we consider the dislocation model of intercrystalline boundaries. The main point is that there is never a free energy minimum at a moderate concentration of dislocations. There is no moderate concentration corresponding to thermodynamic equilibrium, as there is in the case of point defects.

8. – Interaction between dislocations and point defects.

In analogy with the precipitation of impurities in a plane, occurring e.g. in duralumin, NABARRO was led to think of a condensation of vacancies in a monoatomic layer, forming a dislocation ring. This situation is pictured in Fig. 12, for the case of a *sc* lattice.

Vacancies can be added, widening the dislocation ring. The dislocation ring can easily glide from one lattice plane to another. However, no metal crystallizes in the *sc* structure; for this reason, let us study the situation in the *fcc* lattice. The close-packed planes in this lattice are the (111) planes: their stacking is of the type ΔΔΔ as illustrated in Fig. 13.

The two permissible modes of stacking designated by symbols Δ and ∇ lead to three possible projected positions of the close-packed planes, designated by *A*, *B*, *C*. The cyclic order *ABC* corresponds the to Δ stacking, cyclic order *CBA* to ∇ stacking. Hexagonal close packing corresponds to Δ∇ stacking in alternation, or *ABAB*. Let us suppose vacancies condense in an *A* plane:

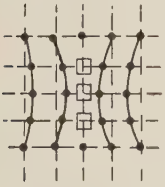


Fig. 12.

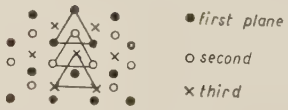


Fig. 13.

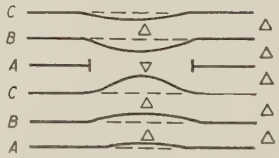


Fig. 14.

the planes above and below this plane will collapse, and we will have a plane *B* above in contact with a plane *C* below, namely a ∇ stacking inter-

posed in the Δ sequence. This is called a stacking fault. The dislocation ring surrounding it is unable to glide. If the dislocation were to glide, it would bring the edge of a B plane to the edge of an A plane, and since B and A are different planes of atoms they will not fit. This glide would create a surface of high energy, of the same order of magnitude as the energy of a grain boundary in a typical polycrystalline metal. The stacking fault itself is of finite energy, of the order ten times smaller than the grain boundary energy, as we infer from the measurable energy of twin boundaries, having a closely related configuration, in *fcc* metals. The twin boundary is a place where the regular $\Delta\Delta\Delta$ sequence changes over to a regular $\nabla\nabla\nabla$ sequence, representing the *fcc* lattice in another orientation.

The electron-microscopic film made by HIRSCH, WHELAN and BOLLMANN shows dislocations in motion in thin foils of aluminium and in 18—8 stainless steel—both f.c.c. metals. The foils are about 2 000 Å thick, prepared by rolling and electro-polishing. Dislocations are visible because of modified diffraction in the strongly strained region of the crystal near to the dislocation line—much of the image contrast in these crystalline films is attributable to Bragg reflection. The dislocations lie systematically in (111) planes, in which they have lowest energy. They are therefore usually oblique to the plane of the foil and are seen as short lines in projection. The path along which the dislocation has recently travelled in the metal is also visible, because, in effect, trailing dislocation lines are trapped against the oxide films at the surface. The effect of this pair of trailing dislocation lines is to change the crystal orientation slightly between them, making a visible darker or lighter band. This effect fades, presumably by relaxation in the oxide layer.

The motions seen must not be regarded as completely typical of dislocation motions in bulk metal. There is not room in thin foils for the dislocations to carry out the complex evolutions which occur in macroscopic specimens. A consequence is that metal foils thinner than 1 μm are anomalously strong. Likewise metal « whiskers » with a thickness of the order of 1 μm are anomalously strong. This is not necessarily *always* to be attributed to the absence of dislocations in the whiskers: restrictions on their motion preventing the dislocations from multiplying may sometimes suffice.

The phenomena shown in this film were: « shunting » motion of dislocations, revealing the repulsive forces between dislocations of like sign in or near the same slip plane; « cross-slip » of dislocations, whereby a dislocation prevented in some way from continuing to glide in one [111] plane diverts into another: a dislocation for the slip direction $[1\bar{1}0]$, when it assumes the screw orientation is simultaneously parallel to the alternative slip planes (111) and $(11\bar{1})$; and finally, the bowing-out of dislocation loops under stress, ultimately moving forward after passing a critical value of curvature.

HIRSCH *et al.* have also succeeded in showing the forced expansion of dislocations into wide strips of stacking fault.

The source of the stress under which these dislocations move is not under the direct control of the experimenter. In part it is the heating of the specimen in the electron beam. Probably it is also in part due to « radiation damage » in the oxide film.

9. – Burgers vector.

The Burgers vector of a dislocation may be defined in the following way: consider a closed circuit enclosing the dislocation line in the real crystal, and let us map this circuit on an ideal crystal, going stepwise from one corresponding lattice point to the next one. This mapping is a locally continuable process (by « tetrahedrulation » from atom to atom) if, and only if, the distortion of the real crystal is not excessively large. Regions where it is possible without ambiguity are called « good » — other regions are called « bad ». The circuit made continuously in good crystal may enclose a bad region, and then the corresponding circuit in the ideal crystal may not be closed: the closure vector is called the Burgers vector $\boldsymbol{b}$. It is necessarily a lattice translation.

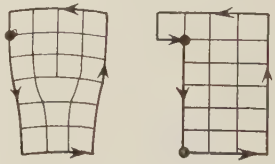


Fig. 15.

It is easily verified that the Burgers vector so defined is independent of the starting point. A vector $\boldsymbol{b}'$ defined by mapping $\boldsymbol{b}$ back into the real crystal is approximately, but not exactly, invariant. The value of $\boldsymbol{b}'$ at the dislocation is to be defined by a limiting process, as the mean of its value at surrounding points. It is sometimes, but not always, an unnecessary subtlety to draw this distinction.

For an edge dislocation the Burgers vector is perpendicular to the dislocation line. For a screw dislocation it is parallel to the dislocation line. Circuits can be quasi-continuously displaced through good regions, without change of closure vector, hence one has the continuity theorem: the dislocation line cannot terminate in the interior of the crystal.

If one has two regions containing dislocations in the crystal, one finds easily that the resultant Burgers vector is the sum of the Burgers vectors of the individual dislocations (see Fig. 16). From this it follows that junctions between dislocation lines are allowed, provided, $\boldsymbol{b}_3 = \boldsymbol{b}_1 + \boldsymbol{b}_2$, as is seen by considering the two circuits drawn. This condition may be written $\sum \boldsymbol{b} = 0$, if one looks along all the dislocations from the junction, when choosing the sense in which to take a circuit.

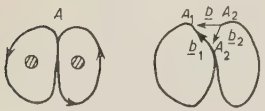


Fig. 16.

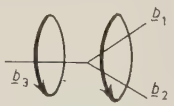


Fig. 17.

Let us consider again the situation in a f.c.c. lattice, where in part of the plane a ∇ stacking is interposed in the Δ sequence. Now it is impossible to find a circuit for one of the dislocations passing entirely through good crystal. The circuit must start from the stacking fault as shown. Supposing the distance between the ∇ stacked plane to be equal to the ordinary distance between (111) planes, one finds as a Burgers vector $-\frac{1}{3}(1, 1, 1)\,L$ and $\frac{1}{3}(111)\,R$ for the left and right sides of the diagram. The values $\frac{1}{3}(111)\,L$ and $-\frac{1}{3}(111)\,R$ are not allowed, when there is only one ∇ fault, but can occur for a double fault, which might be formed by an aggregation of interstitial atoms.

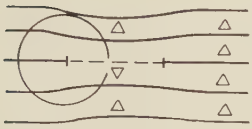


Fig. 18.

Fig. 19.

10. – Extended dislocations.

When part of a dislocation line having Burgers vector $\frac{1}{2}[1\bar{1}0]$ lies in a (111) plane, that part may dissociate into two glissile half dislocations with Burgers vector $\frac{1}{6}[1\bar{2}1]\,L$ and $\frac{1}{6}[2\bar{1}1]\,R$. These dislocations repel each other and, as they separate, a sheet of stacking fault is formed in the slip plane between them. The energy of this fault prevents them from separating too far. This gives a constant attraction, balanced against the repulsion, proportional to the scalar product of the Burgers vectors of the two partial dislocations, and inversely proportional to the distance between them. Equilibrium widths of 10 or 15 atomic spacings are estimated for copper, but only about 2 atomic spacings for aluminium, which has a higher stacking fault energy. The difference in width of extended dislocations in various metals gives rise to important differences in some experimentally observable properties.

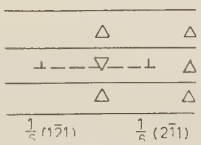


Fig. 20.

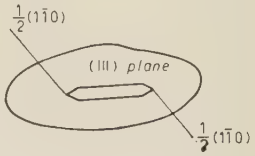


Fig. 21.

It may be mentioned that the technique of HIRSCH *et al.* will permit a more direct estimate of the stacking fault energy, whereas until now calculations have been based upon the assumption that this energy is comparable with grain boundary energy.

11. – Adding of dislocations.

A pair of dislocations will add if the new dislocation has smaller energy than the original pair. From this principle, one finds, if the isotropic elasticity

theory is applied, if the difference between screw and edge is neglected, and if the core energy is neglected, dislocations will add if their Burgers vectors make an angle smaller than 90° as shown $\overrightarrow{\swarrow}$ (uniting dislocations).

12. – Cottrell-Lomer barrier.

Let us consider two extended dislocations $\frac{1}{2}[10\bar{1}]$ and $\frac{1}{2}[011]$ sliding on different intersecting planes. The partials $\frac{1}{6}[2\bar{1}\bar{1}]$ and $\frac{1}{6}[\bar{1}21]$ will interact at the junction to form a partial $\frac{1}{6}[110]$.

This partial, together with the other partials $\frac{1}{6}[11\dot{2}]$ and $\frac{1}{6}[112]$, repel each other. Thus a bent ribbon of extended dislocations has formed; which is unable

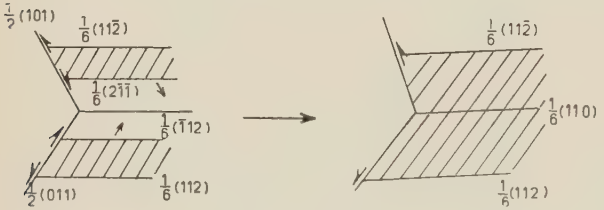


Fig. 22.

Fig. 23.

to glide in any direction. This situation may result in a pile up of dislocations. Such pile ups have been observed by JACQUET using an etching technique.

13. – Intercrystalline boundaries.

The figure shows the simplest type of tilt boundary in which two crystals of simple cubic lattice, slightly misaligned to each other, are joined together. Such a boundary involves a large number of parallel edge dislocations. This configuration is stable in respect of glide motions, as can be easily seen from the distribution of shear stress around an edge dislocation, described in Section **4**.

So long as the angle of misalignment α is small, so that the spacing d between the dislocations (given by $b/d = \alpha$) is large, the dislocations are distinct individual dislocations. Their presence may be revealed by etching or other techniques. The dependence of boundary energy on misalignment may be calculated from the dislocation model, in satisfactory agreement with observation; the observed values may then be used to estimate the core energy of

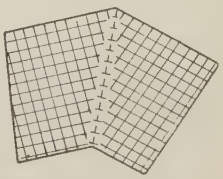


Fig. 24.

the dislocations, as mentioned earlier. The agreement is actually satisfactory up to higher dislocation densities than we have a right to expect.

In the case of « twist boundaries » in which the misalignment is by rotation about an axis normal to the boundary plane, the corresponding dislocation system is a crossed grid of screw dislocations: or, at surfaces of trigonal symmetry an equivalent hexagonal network made up of screw dislocations of three types.

In the general case of misalignment by a rotation α about an axis defined by the unit vector $\boldsymbol{l}$, the dislocation content of the boundary is such that if $\boldsymbol{r}$ is the vector connecting two distant points in the boundary (which is not necessarily plane), any line in the boundary connecting these two points crosses dislocation lines having the same sum of Burgers vectors $\sum \boldsymbol{b}$, and

$$\sum \boldsymbol{b} = \boldsymbol{r} \times 2\boldsymbol{l} \sin \tfrac{1}{2}\alpha .$$

The proof of this formula is easily given by constructing a Burgers circuit intersecting the boundary at the two specified points. So long as the misalignment is small, and the dislocation density accordingly low, this may be done in such a way that the circuit passes continuously through « good » crystal. This is no longer so when the misalignment is large, and alternative specifications of the misalignment and dislocation content then arise. A cubic crystal may be rotated into parallelism with another in 24 different ways. If we always choose the rotation of smallest angle, the largest value the angle can take is about 63° and then there are four alternative equal rotations. As a simple extreme example, one may consider the symmetrical tilt boundary about a cube axis when the angle α goes to 90°. This can either be described as having an atomic density of dislocations or no dislocations at all.

Fig. 25.

It is this kind of ambiguity which destroys the usefulness of describing a liquid as a crystal full of dislocations — the very high density of dislocations has no unique specification.

Photographs were projected showing examples of:

1) Parallel edge dislocations which lie along crystallographic directions.

2) Hexagonal grids of screw dislocations (Fig. 1), as discussed by FRANK in *Report of the Bristol Conference on Defects in Crystalline Solids*, p. 159 (1955).

3) Grids of dislocations in silver halides decorated by Mitchell by deposition of metallic silver, after short exposure to light. The patterns shown demonstrate the cell structure of the crystal, the blocks being of the order of 10 μm in size. The surfaces between these blocks are various kinds of grids of dislocations, most commonly either parallel lines or hexagonal nets. This

was the first real evidence of the mosaic structure, suggested long ago to account for X-ray diffraction intensities. The description of real crystals in terms of dislocations is a more flexible one than in terms of mosaic walls. Sometimes there are crystal blocks separated by walls of dislocations. Sometimes there is only one dislocation per wall, and it is better then to speak of a three-dimensional dislocation network and not to call it a mosaic structure.

4) Parallel arrays of dislocation lines in alkali halides, decorated with gold, by heating the alkali halides in the presence of gold halide.

5) Dislocation grids at the interface between crystals of different lattice spacing. These are the moiré patterns of PASHLEY, MENTER and BASSETT (*Nature*, 179, 752 (1957)), obtained in the electron-microscopy of superposed films of gold and palladium. There are three equivalent ways of interpreting these patterns:

(*a*) As a simple superposition pattern such as one sees by looking through two layers of silk, or two fences, with bands of light and dark where the two fine patterns come in and out of step with each other — having a spacing about 10 times that of the fine patterns when the spacings differ by 10% as is the case with gold and palladium. It may then be shown by trial that a dislocation in one of the fine patterns, unless matched by another one in the other, makes a corresponding dislocation seen on a larger scale in the moiré pattern.

(*b*) As the result of successive diffraction by the two layers. Every imaging process can be considered in terms of diffraction, and ought to be so considered when the resolution is in question. In the present case the origin of the simple, undislocated, moiré patterns is readily interpreted in this way. The detailed calculation would be more difficult for the case with a dislocation in one of the films, but an application of Parseval's theorem reduces the problem to interpretation (*a*) without requiring calculation in detail.

(*c*) We may consider that we are seeing by electron-microscopy (apart from the unavoidable imperfection of imaging due to working near to the resolution limit) the actual grid of dislocations which must exist between the two phases. Various whole numbers of moiré bands may correspond to one interfacial dislocation. In a commonly occurring case we can show that each moiré band represents one partial dislocation at the interface. In this interpretation the appearance of dislocations in the moiré pattern is a consequence of the continuity theorem, since each dislocation that emerges through one of the foils without passing through the other must be drawn from the interfacial grid. This way of interpreting the patterns is particularly valuable when the two films are in different orien-

tations instead of, or as well as, having different spacing — in these cases the orientation of the dislocation seen in the moiré pattern is not the same as that in one of the foils.

14. – Forces on dislocations.

Consider a dislocation moving in a crystal and passing between two initially adjacent atoms, P, Q, so as to enter a Burgers circuit constructed through these points. As it does so, the closure vector of the mapped circuit in ideal crystal changes by $\boldsymbol{b}$. But its only alteration is between the map points of P and Q. These points have therefore suffered a relative displacement $\boldsymbol{b}'$. The motions we consider in this argument need not be restricted to motions in the glide plane. The stress in a body is measured by the stress tensor $P = P_{ij}$, so defined that $|\mathrm{d}\boldsymbol{a}|$ times the traction on an element of area $\mathrm{d}\boldsymbol{a}$ (i.e. the force required to be applied at the faces of this element of area to prevent displacement if a cut is made at it) is $P\,\mathrm{d}\boldsymbol{a} = \sum_j P_{ij}\,\mathrm{d}a_j$.

Then the work done by the elastic field when a dislocation line element $\mathrm{d}\boldsymbol{s}$ sweeps the element of area $\mathrm{d}\boldsymbol{a}$ is

$$(P\,\mathrm{d}\boldsymbol{a})\cdot\boldsymbol{b}' = \sum_i \sum_j P_{ij}\,da_i b'_j = (P\boldsymbol{b}')\cdot\mathrm{d}\boldsymbol{a}\,.$$

Dividing by $|\mathrm{d}\boldsymbol{a}|$ gives the force due to the elastic stress field tending to move the dislocation line element $\mathrm{d}\boldsymbol{s}$ in that direction in which it sweeps the area $\mathrm{d}\boldsymbol{a}$. The force per unit length may be seen either as the scalar product of the traction on the swept area with the Burgers vector, or $|\boldsymbol{b}|$ times the scalar product of traction on a plane normal to the Burgers vector with the unit vector normal to the swept area. Considering all the planes in which the line element $\mathrm{d}\boldsymbol{s}$ can move, we see that this corresponds to the existence of a force $(P\boldsymbol{b}')\times\mathrm{d}\boldsymbol{s}$ on each element $\mathrm{d}\boldsymbol{s}$ of the dislocation line.

15. – The force for glide motion.

The glide motion of a dislocation is one causing no change of volume of the crystal. It is motion in the surface which contains both the dislocation line and its Burgers vector, which is unique except for a screw dislocation. Commonly the glide surface is a plane. The tangential component of traction on a plane is called the shear stress on that plane. Then if $\boldsymbol{\sigma}$ is the shear stress on the glide plane, we have from the foregoing formula that $(\boldsymbol{\sigma}\cdot\boldsymbol{b}')$ measures the force per unit length on the dislocation line tending to move it in its glide plane. This is a uniform force normal to the dislocation line, independent of

dislocation orientation. In conjunction with the line-tension approximation, it gives us a close analogy, in two dimensions instead of three, between the equilibria of dislocations under stress and soap bubbles under one-sided pressure.

16. – Forces for climb motion.

The force $(P\boldsymbol{b}')\times \mathrm{d}\boldsymbol{\sigma}$ also contains in general a component normal to the glide plane of the dislocation. It is readily visualized as the force tending to « squeeze out » the inserted half-plane. Such motion, however, involves a volume change, which can only be brought about by creation or absorption of interstitial atoms or vacancies at the dislocation. Such motions require diffusion processes, are relatively slow, and do not occur at low temperature. They are called climb motions in distinction from glide. If the concentration of point defects is not in equilibrium, this gives another force, the osmotic force, on the dislocation, in addition to that due to elastic stresses. The osmotic force due to a concentration of vacancies which exceeds the equilibrium concentration by a factor α is

$$(\boldsymbol{b}'\times \mathrm{d}\boldsymbol{s}/v')kT\ln\alpha\,,$$

where v' is the lattice volume per atom.

I must now mention briefly a number of topics which I have no time to deal with in detail but are too interesting to leave unmentioned altogether.

17. – Pile-up equilibrium.

In a slip-band we have a number of dislocations in the same or nearly the same glide-plane, each acted upon by the same force from the applied stress, and repelling each other with a force inversely proportional to distance. If the leading one is held up by a barrier of some sort, we have the problem of finding their equilibrium distribution. This looks like being a very difficult problem for a large number of dislocations, but is actually very simple: the solution was given by STIELTJES in 1885 (without reference to dislocations, of course). The theory was applied to dislocations by ESHELBY, FRANK and NABARRO. For a shorter account see FRANK, Pittsburgh Conference 1950, O.N.R.

18. – The origin of dislocations.

This is something we do not know enough about. There are probably mechanisms of dislocation production associated with inhomogeneous distribution of impurities. For pure materials the principal process preventing us from

obtaining large dislocation-free specimens is probably the formation of dislocations by vacancy aggregation, both in the moving temperature gradient while the crystal is being grown, and also during the process of cooling down after growing or annealing. For reasonable cooling times it is only in very thin specimens (especially metal whiskers) that there is time for vacancies to diffuse to the free surface. The interesting and difficult problem is to decide whether they should then aggregate into dislocation loops or open cavities. The problem is difficult because the fate of the aggregate is probably decided when it consists of 5 or 6 vacancies — too small a number to be considered macroscopically, and a large number to be considered discretely. Some discussion of the matter will be found in the report of the 1956 Lake Placid Conference (shortly to be published) starting with a contribution by J. C. FISHER.

19. – Kinking theory.

Fig. 26.

When a pair of boundaries of opposite sign are close together, there must exist a shear stress of the form shown to keep the rows of dislocations apart. The region *A* must be heavily stressed. The theory of this process, which is important in twinning and martensitic transformations, has been discussed by FRANK and STROH (*Proc. Phys. Soc.*, B **65**, 811 (1952)). It is further discussed in the book of FRIEDEL.

20. – Adsorption and precipitation at dislocations.

The first application of this idea was by COTTRELL who explained various phenomena in mild steel by the adsorption of carbon atoms at dislocations. We now have many examples of the precipitation of impurities at dislocations, making the dislocations visible. This is attributable to three concurrent effects: the adsorption of impurities at dislocations, the enhanced diffusivity along dislocation lines and reduced strain energy of nucleation in an already severely strained lattice.

21. – Interaction of electrons with dislocations.

This is a topic deserving at least a whole lecture to itself, but I shall mention only the two basic ideas of the subject. The publications of READ deal specifically with dislocations in covalently bonded crystals, in which we suppose there are what SHOCKLEY calls « dangling bonds » along the dislocation. These

are regarded as essentially equivalent to impurity centres with some important distinction arising from their relatively dense distribution along a line. KAWAMURA has suggested that in a more general class of crystals, e.g. KCl, the deformation potential may suffice to give line-bound states for electrons.

22. – The work-hardening curve.

This was the subject of TAYLOR's original paper in 1934, and, together with metal creep, has been the subject of more studies in dislocation theory than anything else. I have purposely omitted it from this course, partly because I wanted to concentrate on points which may connect with wider branches of solid state theory, partly because very full discussions of the matter are available in published lectures by MOTT, in the books of COTTRELL and FRIEDEL, and the Handbuch article by SEEGER, and partly because the full truth of this complex matter has yet to be said.

23. – Jogs in dislocations.

A place on the dislocation line where it passes from one glide plane to an adjacent parallel one is called a jog. If the dislocation is non-screw, then we can imagine that it was formed by insertion of a layer of atoms into a cut in the crystal. The edge of this layer is the dislocation line, and is stepped wherever there is a jog in the dislocation. To describe a given dislocation, the choice of cut and inserted layer is not unique, except as to the position of its edge.

Fig. 27.

For an example let us consider an ionic crystal with NaCl structure. Three equivalent choices for the inserted layer to make a dislocation having Burgers vector $\boldsymbol{b} = (a/2)[110]$ are shown in Fig. 28.

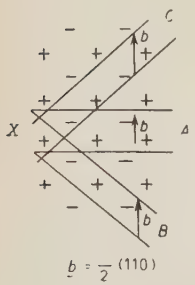


Fig. 28.

The layers shown, chosen from many other possibilities, contain two extra (110) planes in A, or one extra (100) or (010) plane at B or C. In each case the dislocation lies at the edge of the inserted layer, along [001] at X in Fig. 28. It is an edge dislocation for the slip plane $(1\bar{1}0)$. In each case, the faces of the cut must be separated along the same direction and by the same amount along this direction, $\boldsymbol{b} = (a/2)[110]$. In each case, the contents of the layer have the same projection on to the plane normal to $\boldsymbol{b}$ (Fig. 29). Encircled ions in this figure are at an odd, the remainder at an even number of (110) inter-

planar spacings above the projection plane. In this figure the dislocation lies along AB. Suppose now that this dislocation is intersected by another dislocation whose Burgers vector $\boldsymbol{b}_2$ may have any of twelve different orientations

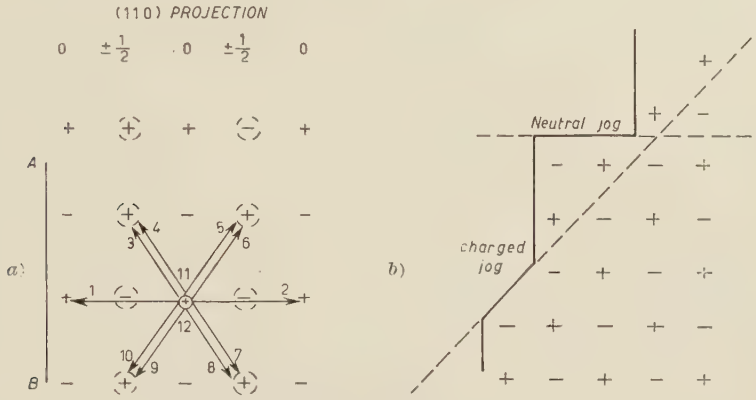


Fig. 29.

represented by the ten arrows on Fig. 29, together with two directions perpendicular to the plane of the paper.

The intersecting dislocation makes a relative displacement $\boldsymbol{b}_2$ of the two portions of crystal either side of its glide plane. As it does this to the inserted layer also, it generally makes a jog in the first dislocation. Except at the edge, the place where the glide plane cuts the inserted layer depends on where we imagine this to be. It is advantageous to imagine this as one of the single planes (100) or (010), so that this is a straight line.

None of the possibilities for $\boldsymbol{b}_2$ is parallel to the line of the first dislocation, so that a jog is produced in every case. Numbers (11) and (12), with $\boldsymbol{b}_2$ parallel to $\boldsymbol{b}_1$, make a « slip-jog », which can disappear again by glide. Numbers (1) and (2) make neutral jogs, with both a positive and a negative ion at the step (Fig. 30). The remaining eight cases, numbers (4) to (10), make a charged jog, together with a jog component which can glide out again. The charged jog carries an effective charge of $+\frac{1}{2}e$ or $-\frac{1}{2}e$ according as the ion at the step is positive or negative.

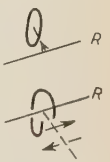


Fig. 30.

Our conclusion here differs from a widely quoted statement by SEEGER, that it is not geometrically possible to form any but neutral jogs by intersection, and that formation of charged jogs requires an additional diffusion process.

Experimental evidence bearing on the presence of charged jogs may be found in experiments by STEPANOW and by NOWICK, showing charge transfers accompanying plastic deformation in NaCl, in addition to the temporarily enhanced conductivity discovered by GYULAI.

24. – Intersecting screw dislocations.

Consider a closed circuit into which a right-handed screw dislocation is moved. The circuit is then opened as shown: thus a clockwise simple shear is produced.

Consider now the intersection of two screw dislocations. Two intersecting right screws produce overlapping material (a line of interstitials). The same applies for two left screws — a mirror image of the previous case. If one screw is right and the other left a line of vacancies is formed. COTTRELL has pointed out that if the intersecting dislocations are activated by the same stress there is a greater probability of the interstitial-forming mechanism occurring. One must, however, enquire whether these processes occur at all.

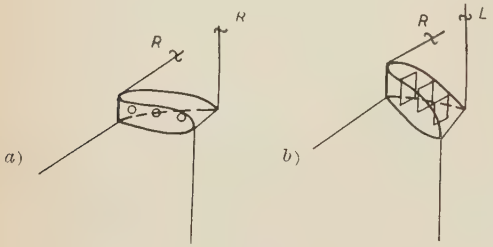


Fig. 31.

Consider an expanding dislocation loop acquiring jogs by intersection with other dislocations. The jog is a little piece of dislocation which can glide in the directions shown. FRIEDEL suggests that there is only a narrow range of orientation of the dislocation from which the jog does not glide away to the edge position, where it can move with the rest of the dislocation without making any vacancies or interstitials; namely, that range of orientation in which the jog would need to move faster than sound in order to do this. I doubt this

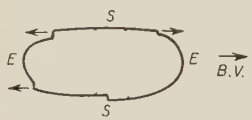


Fig. 32.

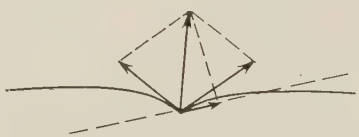


Fig. 33. – Glide force on a jog.

conclusion. Let us consider the force which makes the jog glide. Apart from a component directly due to the applied stress acting on the jog, which is equally likely to assist or oppose its motion, this must come from the line tension of the two portions of dislocation line to right and left of it, making an angle with each other.

Now, I enquire if there is any stationary shape in which the dislocation line can move, drawing the jog along with it. I think there cannot be, because this requires the dislocation line to be moving faster at a place where it has

convex curvature, faster than at a place where it is relatively straight. This must be wrong, and I believe that when the main part of the dislocation line continues for a while in steady motion, its jog must hang back more and more (Fig. 35). This produces a pair of edge dislocations, of opposite sign, attracting each other, in glide planes one atomic layer apart. This is just equivalent to a line of vacancies (Fig. 36) or a line of interstitial atoms for a jog of op-

Fig. 34.

Fig. 35.

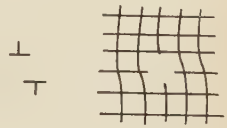


Fig. 36.

posite sign. If the temperature is not too low, such a line will be expected to evaporate into independent point defects.

Other suggestions which have been made to account for the Gyulai effect depend on producing the corresponding configuration by accidental meeting of a pair of edge dislocations of independent origin, one (or perhaps two or three) atom planes apart: or, alternatively, depend on the localized energy release when such a pair meet in the same plane and annihilate each other.

25. – Spiral processes.

Under this term I will group together several different processes which are geometrically related to each other. The first is crystal growth. The Volmer theory of crystal growth, which is the right theory for undislocated crystals, does not fit the experimental facts. It assumes independent nucleation of new monolayers on the crystal surface, and spreading of these layers by addition of atoms along the step which is the edge of the layer. It leads to a nucleation rate which is very sensitive to supersaturation. In any crystal growing from a solution, the supersaturation is highest at the corners, because of diffusion. New layers ought therefore to nucleate at the corners before there is time for them to spread across the face. Crystals growing by this mechanism will therefore always be dendritic. The supersaturation required for this mechanism is also much higher than is observed to be necessary. The reason is that crystals containing screw dislocations have permanent steps in the surface. They do not require nucleation of new layers, because the crystal consists of one helicoidal layer only. Atoms continuously add themselves to the step, making it rotate around its fixed point, the end of the dislocation. In doing so it does not remain straight, like the hand of a clock, because there is no reason why

its speed should increase proportionally to its distance from the centre. Its speed is much more nearly constant along its length, and it therefore winds itself up into a spiral. Near the centre of the spiral it acquires the same curvature as the critical nucleus of Volmer's theory, which is in unstable equilibrium with the supersaturation.

Suppose now we have a right and left screw emerging close together at the surface. If they are too close together for the equilibrium curvature (so that the Volmer critical nucleus could not pass between them) no growth occurs. The curved step is in stable equilibrium with the supersaturation. Increased supersaturation increases the curvature, and beyond a critical value growth proceeds. Spirals commence to form around both dislocations ends, but two steps meeting face to face form a level surface. Then we have a closed step ring around the dislocation pair, and a short length of step connecting them, which can repeat the process. This provides the basis of a successful quantitative theory of crystal growth rates. The corresponding surface configurations have now been observed on a great variety of crystals.

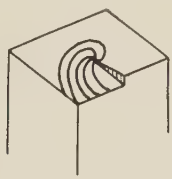


Fig. 37.

Fig. 38.

Fig. 39.

Exactly the same geometry applies to certain kinds of dislocation motion in the interior of crystals. Under shear stress, a dislocation line held at two points bows out in its glide plane like a two-dimensional soap-bubble. Increased stress brings it to a maximum curvature after which it can expand continuously with decreasing curvature, repeating the configuration of Fig. 39. This (the so-called Frank-Read source mechanism) is believed to be one of the most important processes by which dislocations are multiplied in plastic deformation. The first unambiguous observation of it was made by W. Dash in silicon last year.

A dislocation line can similarly be bowed out by the osmotic force in its climb plane, and perform the evolutions of Fig. 39 in this plane, as first suggested by Bardeen and Herring. However, this case has an interesting variant, the idea of which is due to Dekeyser, Amelinckx, Bontinck and Seitz. A dislocation intermediate between edge and screw should undergo these evolutions, as seen in projecton along the Burgers vector, except that where they meet and cancel each other in the other cases, they are now actually far apart, and pass each other by. The process can then develop a helix of

many turns. This is one way in which the helical dislocations found by the GHENT workers can be explained. I think there is perhaps another explanation, chiefly because their helices, first observed in CaF_2, with evaporated silver on the surface, heated in hydrogen, are so accurately oriented along screw dislocation directions.

A pure screw dislocation is not able to climb, but if on account of thermal oscillations a segment deviates from the straight dislocation line, edge components E appear and climb can occur. They are of opposite sign and will climb in opposite directions. Thus the straight screw dislocation can be transformed into a helix. A related process has been osserved by MITCHELL in NaCl decorated with gold. Microscopic cavities in the crystal are seen to be accompanied by a series of dislocation rings in opposite [110] directions. It is thought that decomposition of gold chloride simultaneously produces shear stresses by liberating chlorine gas at high pressure in the cavity, so initiating slip, and also makes a non-equilibrium concentration of point defects to cause climb making dislocation loops in the manner envisaged by BARDEEN and HERRING except that they do not remain in the same plane but are driven away by the inhomogeneous shear stresses.

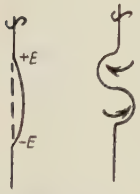


Fig. 40.

Fig. 41.

BIBLIOGRAPHY

A. H. COTTRELL: *Dislocations and Plastic Flow in Crystals* (1953).
W. T. READ: *Dislocations in Crystals* (1953).
J. FRIEDEL: *Les dislocations* (1956).
A. SEEGER: *Handbuch der Physik*, Vol. VII/1, p. 383 (1955).
W. DEKEYSER and S. AMELINCKX: *Les dislocations et la croissance des cristaux* (1955).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Point Imperfections in Solids (*).

F. SEITZ
University of Illinois - Urbana, Ill.

I.

General defects.

The notion of a perfect crystal is an idealization; what we actually have at best at the present time are crystals with a small, controlled number of imperfections. The following classification of the principal crystal imperfections was given by the author a number of years ago, and will be used in the discussion:

1) Phonons,
2) Free electrons and holes,
3) Excitons,
4) Vacancies and interstitials,
5) Foreign atoms,
6) Dislocations.

There are two important remarks to be made:

a) There is only a finite number of imperfections. Each one has been introduced initially to explain particular phenomena.

b) Imperfections interact. A large number of effects arise through their interactions.

We will now consider the main properties of these imperfections.

—

(*) The text is based on Lecture Notes collected by a group of students (W. GEBHARDT, E. GERMAGNOLI and R. HOWARD) and revised by the author.

1. – Phonons.

A phonon is associated with unit quantum excitation of one of the normal modes of vibration of an ideal crystal. One characterizes these vibrations by a wave vector $\boldsymbol{k}$, by a polarization vector $\boldsymbol{f}_\alpha$ and by the frequency $\omega_\alpha(\boldsymbol{k})$, where the index α denotes the direction of polarization. In the simplest case the modes of vibration correspond to two transverse and one longitudinal mode but in general one does not have this breakdown. The allowed energy levels of a vibrating crystal correspond to having 0, 1, 2, ... quanta of energy $\hbar\omega_\alpha(\boldsymbol{k})$ in each vibration of wave vector $\boldsymbol{k}_\alpha$. One can speak in this way of 0, 1, 2, ... phonons in the state $\boldsymbol{k}_\alpha$.

The simplest way to introduce phonons of a given frequency into a crystal is to link the specimen with a piezo-oscillator. One is restricted in this way to oscillations in the range from sonic modes to about 100 MHz while the range of actual interest extends to 10^{13} Hz. If one connects the crystal to a thermostat, one gets an equilibrium distribution of phonons corresponding to the temperature of the thermostat. If $T > \Theta$, the Debye temperature, all the modes will be stimulated.

2. – Free electrons and holes.

At low temperatures, an insulating crystal does not display appreciable ohmic conductivity. If however one excites electrons thermally, or otherwise, the crystal will behave like an ohmic electronic conductor. To explain this, one can use the following model. One excites an electron from the filled valence band to the empty conduction band, leaving behind a hole. In general an electron at the top of the valence band behaves like a positively charged particle while an electron at the bottom of the conduction band behaves as though it were negatively charged. More precisely the sign of the effective mass of the electron is determined by the sign of the curvature of the energy surface in the neighborhood of the electron. The effective mass will be positive or negative according to whether the curvature is positive or negative. A hole behaves electrodynamically like the electron which is missing from the band.

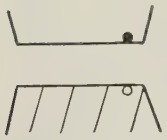


Fig. 1. – The valence and conduction bands of an insulator.

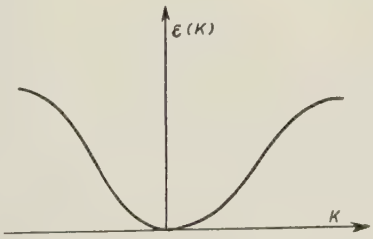


Fig. 2. – Schematic representation of the energy versus wave number curve for electrons in a band.

3. – Excitons.

The nature of free electrons and free holes is quite clear in a typical insulator. One can often produce a controlled number of each defect (n- and p-type semiconductors). One also expects, however, the existence of nonconducting states in which electrons and holes are bound together and behave like a neutral particle. These are commonly designated exciton states.

In crystals having a large index of refraction ($n > 3$) one expects to describe the excitons with the use of the band approximation, the exciton states corresponding to hydrogen-like states of electrons and holes.

In 1931 FRENKEL introduced excitons from another approach. Consider an array of atoms, some of which are excited; FRENKEL considered the quantum mechanical description of the system in the tight binding approximation. One finds a solution corresponding to the wave-like propagation of the excitation and can characterize the solution by a wave number and a polarization vector.

The best experimental evidence supporting this model comes from the properties associated with the absorption peaks of the alkali halides at the beginning of the absorbing ultraviolet region. If one radiates in the first peak one gets no photoconductivity when the crystals are free of color centers. This can be explained if one assumes the existence of excitons. Recently TAFT and PHILLIPS at General Electric Laboratories have observed transition beyond the first peaks, associated with exciton production, corresponding to the ionizing band to band transitions.

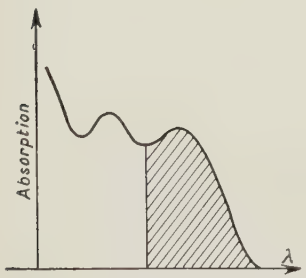


Fig. 3. – The shaded region shows the portion of the absorption spectrum of a typical alkali halide crystal associated with exciton production.

We have thus two ways of looking at excitons, one based on the band theory and the other on the tight binding approximation. We should expect the band theory approximation to be valid for materials with small energy gaps and large indexes of refraction, like silicon and germanium. We should expect the tight binding approximation to be valid for substances with large energy gaps and small indexes of refraction. One would not expect a hydrogenlike spectrum in the case of alkali halides.

Surprisingly, GROSS has found a hydrogenlike absorption spectrum in Cu_2O for which $n \simeq 2.3$. Other evidence of excitons has been given by BALKANSKI and BROSER and by DIEMER and HOOGENSTRAATEN who observed the transmission of energy through large distance in CdS. The crystal is irradiated

at A and charges appear at B under circumstances in which neutral units pass from A to B. Presumably the excitons dissociate into free electrons and holes. The energy is transmitted by means of uncharged particles. Experiments were performed at liquid nitrogen temperature and the observed lifetime of the presumed exciton was about 10^{-3} s.

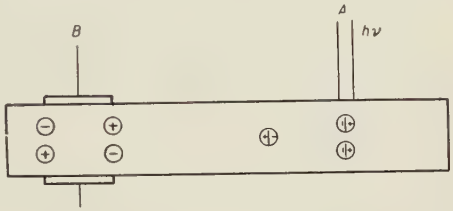


Fig. 4. – The crystal is irradiated at A and charges appear at B under circumstances in which neutral units pass from position A to B. Presumably the excitons dissociate into free electrons and holes.

4. – Vacancies and interstitials.

The key knowledge about vacancies and interstitials has been derived from the study of the electrolytic properties of polar crystals. FRENKEL first observed that the existence of ohmic current in alkali or silver halides required the postulate of mobile charged imperfections. He proposed a model which still bears his name, according to which a cation jumps from a normal site into an interstitial position leaving a vacancy behind. The interstitial ion will move as a positively charged particle while the vacancy will move as a negatively charged one in this model.

Later SCHOTTKY and WAGNER pointed out that in the case of the alkali halides it is more reasonable to expect both anion and cation vacancies as the basic thermally produced imperfections, rather than interstitials and vacancies.

In recent years it has been found experimentally that the silver halides display intrinsic Frenkel disorder while the alkali halides contain anion plus cation vacancies (Schottky disorder).

In the case of metals the nature of disorder was first studied theoretically in 1941. A calculation by HUNTINGTON and SEITZ predicted that the basic imperfections in copper should be vacancies and that the energy of formation of vacancies must be approximately in the range $1.0 \div 1.5$ eV: a recent calculation by FUMI gave a slightly lower value but the two calculations agree within their expected accuracy. The energy to produce interstitials was calculated by HUNTINGTON and SEITZ to be much larger. A large advance was made in recent years when J. S. KOEHLER and co-workers investigated the effects of quenching upon the resistivity of gold wires. Wires were cooled quickly in order to freeze in as many as possible of the vacancies which are in thermal equilibrium at high temperature. Using Bauerle's technique it was found possible to effect such a quench in the range of temperature between 450 °C and 950 °C. The concentration of vacancies changed by about three decades in this range. For the concentration of vacancies within

a metal we have:

$$C_v = A \exp\left[-\frac{\varepsilon}{kT}\right].$$

From the quenching experiments on gold the activation energy ε was found equal to $(0.96 \pm 2\%)$ eV.

General considerations leave very little doubt that vacancies are the imperfections which appear in such experiments. Generally speaking, vacancies disappear at dislocations or cluster when the metal is annealed near room temperature.

Interstitial atoms can be produced in metals by other methods. Most important are Frenkel defects produced when metals are irradiated with energetic particles.

5. – Foreign atoms.

The introduction of foreign atoms can be responsible for very important changes in the properties of some materials; for instance the addition of impurity atoms can transform insulators into semiconductors. Many other examples are possible.

6. – Dislocations.

The concept of the dislocation was introduced to explain the mechanical properties of solids, particularly their high ductility.

In 1929 PRANDTL first introduced dislocations to explain internal friction of plastic origin. If dislocations are present, a crystal can dissipate energy when subjected to cyclic oscillations.

The work of TAYLOR and OROWAN in the next five years formed much of the basis of modern dislocation theory. We will deal with applications of this theory in connection with particular substances. The theory will be developed in some detail by Professor FRANK in his lectures.

II.

Alkali halides electrolysis.

Let us consider the electrolytic conductivity σ observed in an alkali halide crystal. The key experimental observation is that the $\log \sigma$ vs. $1/T$ curves are composed of two distinct, linear parts. There is a straight line portion at high temperatures which deviates as one goes to lower temperature (Fig. 5).

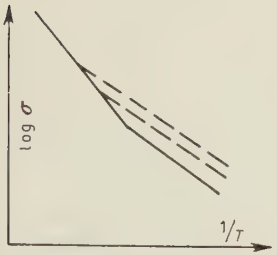


Fig. 5. – Typical Boltzmann plot of the electrolytic conductivity of an alkali halide. The dashed curves correspond to crystals containing additions of divalent halide.

A number of measurements of this effect have been made: the most celebrated being that by LEHFELDT in 1932. The transition between the two portions of the curve occurs at lower temperatures, for purer crystals. Conversely, if one prepares specimens with higher and higher concentrations of impurities, the deviations occur at increasingly higher temperatures.

Another important fact should be mentioned. In the impurity range, i.e. for large impurity concentrations, the positive ions in the alkali halides carry almost all the current. In the intrinsic range, especially at high temperatures, both positive and negative ions conduct. For KCl, for example, about 30% of the conductivity is due to negative ions at temperatures near the melting point.

1. – The Schottky-Wagner model.

This was proposed in the early 1930's to account for the conductivity of the alkali halides and has been verified by subsequent experimental work. The model assumes that in the alkali halides an equal number of positive and negative ion vacancies are present in the pure crystal. Ions move by jumping into vacancies, which are able to wander for large distances.

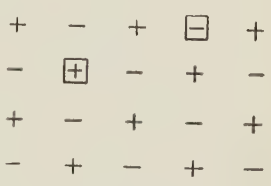


Fig. 6.

2. – Addition of divalent ions.

Divalent ions dissolve substitutionally in the crystal. Because of the extra charge on the ion, this will be associated with a vacancy at low temperature. At higher temperature the vacancy becomes free and, to a first approximation, can be considered to be independent of the divalent ion. A proof of the model was given by the experiments of PICK and WEBER who added $CaCl_2$ to KCl. It was observed that the density of the crystal decreased upon the addition of divalent ions in just the way that would be expected if the divalent ion were added substitutionally and a vacancy were created.

In the diagram (Fig. 7) the top line is what would be expected if densities were additive. The bottom line corresponds to the case in which one Ca^{++} ion replaces two K^{+} ions without causing a change in the lattice parameter.

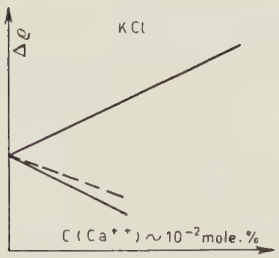


Fig. 7.

The dotted line is what is observed experimentally. The discrepancy between the observed line and the bottom line is presumably a result of shrinking of the lattice parameter as a result of the addition of impurities. This shrinkage is too small to be observed directly in the alkali halides. It can however be observed in other salts, for example those having a CaF_2 type lattice. Trivalent ions like Y^{+++} are highly soluble in these salts. The impurity concentration can be made sufficiently high to produce a change in lattice parameter which can be observed directly. (Divalent ions are soluble in the alkali halides to only about 1 part in 1000). A few years ago MAURER, ETZEL and MAPOTHER studied the effects of adding $CdCl_2$ on the ionic conductivity of NaCl. This type of experiment has been extended considerably since. They found that the isothermal plot of $\Delta\sigma$, the increase in σ above its value for the pure crystals, when plotted against the concentration of added impurity, looks as follows: For low concentrations of impurities, the plot is linear, while for high impurity concentrations there is an observed deviation from linearity. It was proposed by MAURER and co-workers that this effect is a consequence of the association between impurity ions and vacancies. One expects such association since one has a Coulomb attraction between divalent ion and vacancy. MAURER found that the deviation in the conductivity curves could be explained if one assumed an association energy $\varepsilon_a \sim 0.3$ eV. BEAN performed similar experiments with Ca^{++} additions. The conductivity curves for this case showed much more linearity than the corresponding ones for Cd^{++}. This implied an association energy of $\leqslant 0.1$ eV for Ca^{++}. The difference between Cd and Ca is surprising since the two ions are nearly identical, having nearly the same radius and polarizability.

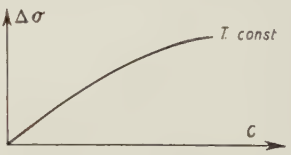


Fig. 8. – Additional conductivity $\Delta\sigma$ as a function of divalent addition.

3. – Theoretical study of association.

The most thorough theoretical investigation was carried out by BASSANI and FUMI. They calculated the following association energies:

	Cd	Ca	Sr
NaCl	0.38 eV	0.38	0.45
KCl	0.32 eV	0.32	0.39

The agreement with the experiments of MAURER and co-workers discussed above is good for Cd^{++} but not for Ca^{++}. However the difference in behavior of Ca^{++} and Cd^{++} impurities is a matter of some controversy, and viewpoints favoring the presence of appreciable association also for Ca^{++} have been expressed by LIDIARD and by HAVEN (see remark 3 at the end of Part II).

MAURER assumed that a vacancy was bound only if it was at a nearest neighbor position to an impurity. If the vacancy was further away it was considered as free i.e. it was considered to be independent of the divalent ion. An extension of the theory was developed by LIDIARD. He made an elegant application of the Debye-Hückel theory of strong electrolytes to take into account impurity-vacancy interactions at distances larger than the nearest neighbor distance. A result of Lidiard's theory is that one gets a higher conductivity than one would predict on the simple theory of MAURER at high temperature and for a given association energy.

4. – Mobility of defects.

One can obtain estimates of the mobility of defects from conductivity experiments on doped crystals. Provided one knows how many free vacancies there are, one can obtain the contribution to the conductivity per vacancy. In order to know the number of free vacancies, one must know the degree of dissociation of the vacancies from the divalent ions.

It is presumably a good approximation to assume that vacancies are free at elevated temperature. Actually one should apply a correction, the correction depending on the model of association one uses. However, when one goes to high temperatures one favors dissociation and the high temperature results are fairly independent of the model one chooses to describe association. Excellent values have been obtained for the mobilities of positive and negative vacancies in NaCl and KCl. Similar measurements could be made for other salts following this approach.

ETZEL and MAURER have found the following

$$\mu_{\boxplus} = \frac{19\,600}{T} \exp\left[-\frac{\varepsilon_r}{kT}\right] \mathrm{cm^2/V\,s}\,,$$

with $\varepsilon_r = 0.8$ eV. At room temperature this would give a jump frequency of about 1 per s. There is an unresolved point which may be mentioned here. The $\log \sigma$ vs $1/T$ plots display a change in slope at low temperature. There is evidence that the impurities may cluster or precipitate out but as yet there is no strictly quantitative treatment of this effect. HAVEN has shown that such condensation is probably very important near 100 °C.

A large amount of additional work of this type has been carried out by ASCHNER and MAURER in recent years and will be published soon.

5. – Dielectric relaxation measurements.

It was noted by BRECKENRIDGE that one should be able to detect impurity-vacancy pairs through dielectric relaxation measurements. If a crystal containing such pairs is placed in an external field having the same frequency as the jump of a vacancy round an impurity ion, dielectric loss should be observed. Breckenridge's experiments were not conclusive, however HAVEN has observed this effect unambiguously. The technique is now regarded as « standard ». One can obtain the activation energy of a vacancy in the neighborhood of a divalent cation by such measurements.

6. – Concentration of intrinsic defects.

Given the mobility of defects and their transport number in the intrinsic range, one can determine the number of defects in the pure crystal from the conductivity isotherms.

The first extensive investigations were carried out by MAURER and coworkers and by the group working at Göttingen with POHL. Other measurements have since substantiated their work but there is still much to be done along these lines.

The following results were found for the densities of intrinsic defects: if f is the mole fraction of Schottky defects for NaCl

$$f = 5.4 \exp\left[-\frac{\varepsilon_F}{2kT}\right],$$

with $\varepsilon_F = 2.02$ eV.

For KCl, KELTING and WITT obtain

$$f = 42 \exp\left[-\frac{\varepsilon_F}{2kT}\right],$$

with $\varepsilon_F = 2.48$ eV.

At the melting point these results give a concentration of Schottky defects of about 1/100 mole %.

Theoretical calculations of the energies of formation of Schottky defects in NaCl and KCl have recently been performed by FUMI and TOSI, and give results in good agreement with experiment.

7. – Diffusion.

Another important way of investigating defects in alkali halides is through measurements of the diffusion of radioactive tracers in the crystal. So far the tracer ions used have been radioactive isotopes of the base atom. However it would be useful to study tracer diffusion of impurity ions in doped crystals. There are difficulties in doing this because sufficiently strong sources of radioactive divalent ions are not available readily.

Using appropriate theory one can infer the conductivity of the crystal from the value of the diffusion coefficient of the tracer. One makes use of the Nernst-Einstein relation

$$\sigma = \frac{D_v c e^2}{kT} ,$$

where c is the concentration of vacancies and D_v the diffusion coefficient of a vacancy. In relating the diffusion coefficient of the tracer ions to that of the vacancy one must take into account the relation between successive jumps of the vacancies.

The need for this correction was first pointed out by BARDEEN and HERRING. The correlation effects are outside present observational accuracy of tracer diffusion measurements in alkali halides; however we shall see that very large correlation effects are observed in the silver halides.

We will conclude with the following brief remarks:

1) An impurity ion should diffuse very rapidly because it has a vacancy attached to it much of the time. The work of ETZEL, MAURER and CHEMLA has borne this out. CHEMLA finds that Zn, Sr and Fe diffuse 1000 times more rapidly than the base ion in NaCl.

2) There has been some work done on the self-diffusion of negative ions in the bromides. KATZ and SCHAMP measured the diffusion coefficient of radioactive Br in NaBr.

3) HAVEN has recently investigated the association of vacancies with Cd and Ca ions with dielectric techniques. It appears from his work that the energy of association of the ions is probably the cause of the discrepancy observed formerly.

4) Experiments have been performed deforming plastically the alkali halide. The conductivity rises when the crystal is deformed and then levels off in time. One concludes that moving dislocations produce vacancies. The effect was first observed by GYULAI and most recently has been investigated by NOWICK.

III.

Alkali halides-color centers.

While in the years before the war the study of color centers in the alkali halides was in the main carried out by POHL and his school at Göttingen and by PRZIBRAM in Wien, research on this subject seems now to be of broad international interest.

1. – Electron centers.

Alkali halides may be colored either by heating in the vapour of alkali metals producing thereby an excess of alkali ions, or by the use of ionizing radiation, such as X-rays. With the use of these procedures one finds a bell shaped absorption band (F-band) in the visible spectrum. The intensity is about $10^{-4} \div 10^{-6}$ times lower than the fundamental absorption of the crystal in the ultraviolet region. By using thin coevaporated films of alkali halides of the type developed by HILSCH, it has also been possible to obtain high F center concentrations, up to $10^{20}/\text{cm}^3$. The oscillator strength of the F band given by several observers has values in the range from $0.7 \div 0.8$.

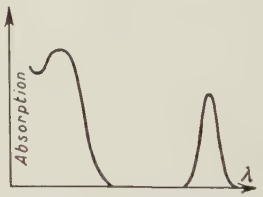


Fig. 9. – The relative positions of the fundamental and F bands.

The plot of log of the concentration of the F centers against log of the vapour pressure of the alkali metal is a straight line with unit slope. This implies the monomolecular process $\text{Na}_{\text{vapor}} \rightleftharpoons \text{Na}_{\text{crystal}}$. The energy needed to dissolve the metal in the solid from the vapor is about 0.1 eV.

The qualitative experimental facts concerning F-centers can be explained by de Boer's model of the F-center, which is supported by a great amount of experimental evidence. In this model a negative halogen ion is replaced by an electron. It is interesting to note that MOLLWO studied CaF_2 to which metallic Ca had been added and showed that the substance colored additively, the coloring process producing absorption peaks similar to the F-centers in the monovalent salts. Density measurements suggest a reaction in which the calcium ions enter substitutionally and the electrons replace fluorine.

The alkali halides containing F-centers possess spin resonance absorption lines, which can be associated with the unpaired electrons in the F centers. HUTCHINSON and KIP and KITTEL observed a broad resonance spectrum. This demonstrates the presence of an interaction of the F electrons with the magnetic moment of the nuclei of the surrounding ions.

PEKAR pointed out from theoretical arguments that colored crystals should

emit fluorescence lines in the near infrared when irradiated in the F band at low temperatures. Although KLICK and his colleagues found very minute fluorescence experimentally, PICK and his co-workers, and DOORN and HAVEN recently obtained luminescent radiation with a quantum efficiency of at least 50%.

Photoconductivity measurements on additively coloured crystals by the Göttingen group show a decrease of efficiency at low temperatures. Fig. 10 illustrates a model of the energy levels of the F center, first described by GURNEY and MOTT, which explains this drop. One obtains photoconductivity only when the electron is raised to the continuum. However the photon lying in the F-band excites the electron only to the first excited, discrete level. The energy difference between the excited state and the continuum is low. At room temperature thermal fluctuations cause further excitation from this level to the ionization states.

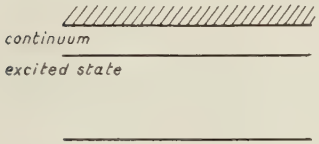


Fig. 10.

Thus most of the excited electrons contribute to the photocurrent at room temperature. On the contrary the excited electrons do not contribute to the photocurrent at sufficiently low temperatures because the thermal energy is not available.

The photoconductivity decreases inversely with the concentration of F centers ($\sigma_I \propto 1/c_F$) at concentrations in the range from $10^{16} \div 10^{19}$ per cm^3. This result indicates that under suitable conditions a second electron can be captured by an F center to form an F' center. The absorption band is shown in the figure. Irradiating in the F-band at appropriate temperature bleaches the F-band and a new F'-band appears. The F' center possesses no excited state. Thus the F' center is ionized by light absorbed in the F' band even at very low temperatures.

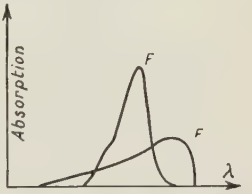


Fig. 11. The F' band arises from F-centres which have captured a second electron.

If one regards the structure of the F band in detail one finds a hump on the tail at the short wavelength side. This is called the K band. Initially it was thought that the K band is the result of excitations to the higher discrete states of F centers. Recent work seems to imply however that the situation is more complicated. The K band may be due to F centers in the neighborhood of dislocations.

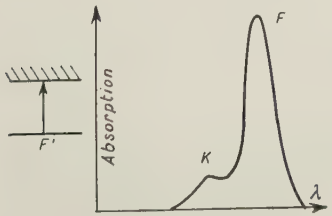


Fig. 12.

While studying the absorption spectrum at the fundamental absorption band of KI crystals colored both additively and by X-irradiation, DELBECQ, PRINGSHEIM and YUSTER discovered two new absorption bands which are termed

the α and β bands. It seems fairly safe to assume that the α band arises from an electron transition between a neighboring halogen ion and the first level of a halogen vacancy. The β band is intimately connected with the F centers. It is reasonable to assume that an electron on the halogen ion is

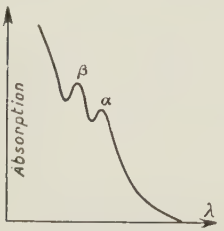


Fig. 13

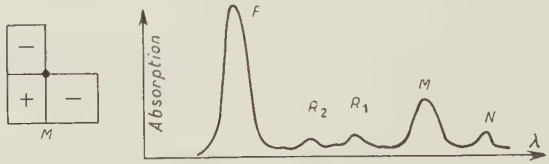


Fig. 14. – Model of the M band. The optical absorption spectrum of a typical alkali halide containing the M band; other bands are shown as well.

excited to the second level of the F center.

When a crystal containing dispersed F centers is irradiated with light in the F band at temperatures near room temperature, a peak near 8000 Å called the M band appears. This absorption is associated with a more complex center which can be formed at sufficiently high temperatures because of the migration of the ions. A model first proposed by the author is shown in the figure. This model was supported by observations by UETA in Japan and by DOORN and HAVEN in the Netherlands. They examined the M centers with polarized light and demonstrated that the center is anisotropic and possesses an axis orientated in the 110 direction. It was also shown that absorption in the F band causes simultaneously absorption in the M center. Thus the M center must be considered to possess an absorption spectrum with two peaks which extend over a relatively large region of the spectrum, one component lying under the F band in KCl.

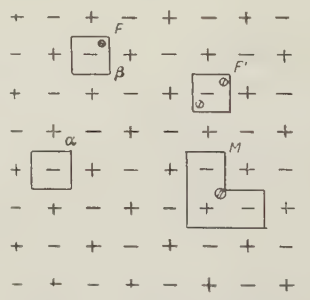


Fig. 15. – Typical centers formed by combining electrons and vacancies.

A more complicated absorption spectrum is obtained after sufficiently long irradiation in the F band. It implies the existence of even more complex centers.

2. – Hole centers.

It is possible to introduce centers analogous to those described above but involving holes instead of electrons. Such color centers absorb in the ultraviolet region of the spectrum and hence have been termed V-centers. In the

case of the bromides and iodides some of the centers, lying farthest to the ultraviolet, can be formed by additive coloration (MOLLWO, 1934), that is by heating the crystals in halogen vapor. Such centers are of bimolecular form and probably correspond to the union of one or two holes with a pair of positive-ion vacancies.

A center which may well be the inverse analogue of the F-center can be formed by irradiating with X-rays or electrons at liquid nitrogen temperatures. This center is called the V_1 center and usually lies relatively near the visible. Although the experimental evidence is not conclusive, it probably consists of a positive-ion vacancy and a hole. It was discovered by DELBECQ, YUSTER and PRINGSHEIM and has been subject to extensive study. It bleaches when the crystals are warmed appreciably above liquid nitroged temperature. DUTTON and MAURER have found that a mobile charge, presumably a hole, is released when the absorption band bleaches. Recently, CHIAROTTI and INCHAUSPÉ have found that the charge released when the V_1 band bleaches can be observed at the corresponding temperature in crystals which have been irradiated with light in the tail of the fundamental band at liquid nitrogen temperatures. The F-band is barely detectable in such crystals and the V_1 band has not been discerned by other means. The current pulse observed is remarkably large and implies that the holes wander until they find electrons and become annihilated.

KÄNZIG has discovered magnetic resonances among the V centers. Thus far, however, the V_1 center has not yielded a detectable resonance, presumably because the resonance spectrum is very broad.

IV.

Silver halides.

1. – Absorption properties.

The principal practical interest in the Ag halides lies in their use in photographic emulsions. We will say very little about this aspect and focus our attention on understanding the fundamental imperfections in the crystal.

Fig. 16 illustrates the general features of the absorption spectrum of a Ag halide. A peak occurs at 2 500 Å in AgCl. Unlike the corresponding spectra for the alkali halides, the Ag spectra exhibit a broad tail extending into the visible region. The first experiments indicated that the absorption curve was not temperature dependent but investigations since 1950 by MILLIMAN and those of GILLEO, TUTIHASI, KOESTER and GIVENS, and OKAMOTO show that at low temperatures there is a sharpening of the peak.

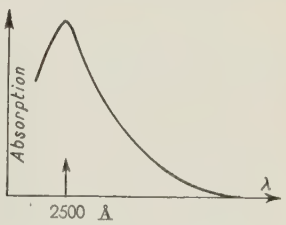


Fig. 16.

The Ag halides show photoconductivity at all temperatures above about 40 °K for radiation in the absorption band. It is of great interest that the quantum efficiency is less than 1 for these crystals.

Brown and Dart at Urbana have extended the older work of Lehfeldt. They observed the quantum efficiency to be about 50% for irradiation in the tail of the band. The efficiency decreases as one irradiates toward the peak of the band.

A possible interpretation of this effect is as follows: the irradiation produces an exciton. In some cases the exciton dissociates producing a free hole and an electron probably in a collision with an imperfection. In other cases the exciton may lose its energy by exciting silver ions. The transition $(4d)^{10} \rightarrow (4d)^9 5s$ in the Ag ion lies in the near U.V. region. One might expect that some of the excitons may lose their energy in this way. In the case of the alkali halides the energy gaps are too large to make such transitions possible.

We turn now to consider the lattice defects present in a pure silver halide crystal. About 20 years ago Jost proposed that the intrinsic defect in the Ag halides is of the Frenkel rather than the Schottky type, i.e. that there are pairs of Ag ion interstitials and vacancies in the pure crystal. Jost was guided by two considerations.

1) The electrolytic current is purely cationic. In the case of the alkali halides there is, in contrast, some anionic contribution.

2) Frenkel disorder is favored in the Ag halides because the Ag^+ ion has a small radius and because of the large Van der Waals energies associated with the high polarizability of the silver ions. The Van der Waals energy between two ions is attractive and varies as r^{-6}; it therefore lowers the energy of interstitials relative to normal ions.

2. – Test of mechanism.

Koch and Wagner in 1936 developed a method for testing this model. Their experimental technique was called in doubt by Mitchell and experiments were repeated by Teltow and Ebert.

The procedure is as follows: one adds various known amounts of divalent impurity to the crystal and then plots the conductivity σ against c. For every divalent ion added a positive ion vacancy must be formed to preserve electroneutrality. The creation of vacancies will depress the concentration of interstitials according to the quasi-chemical reaction

$$Ag^+ \rightleftharpoons \boxed{Ag^+} + Ag_\bullet^+ .$$

If the interstitials are more mobile than the vacancies the conductivity should decrease when impurities are added. What actually was observed was that the conductivity first decreased and then increased. One concludes therefore that the mobility of the interstitials is greater than that of the vacancies but that the vacancies do carry some of the current. Thus one obtains a rise in the conductivity if their concentration is increased sufficiently.

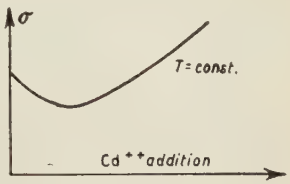


Fig. 17.

It may be mentioned that the analysis of this type of experiment yields the mobility of the two defects and their density in the pure crystal. The density of Frenkel defects is about 1% at the melting point. This is to be compared with the corresponding value of 0.01% for the alkali halides. The mobility and density of defects in Ag halides are known to a factor of about 2. An important question which has received much attention is the degree to which Schottky defects are present at temperatures near the melting point. The technique outlined above loses much of its validity at high temperatures because the mobilities of the defects become nearly equal so one must use other approaches.

Two types of evidence lead one to believe that defects in the Ag halides are predominantly of the Frenkel type even at the melting point.

1) BERRY studied the change in cell dimensions at high temperatures by X-ray measurements. He compared his results with former density measurements made by STRELKOW. The fractional change in the volume of the crystal and the cell are parallel to within the experimental accuracy of 0.1%. The conclusion is that the defect formed preserves the total number of cells. This is true for Frenkel but not for Schottky defects. ZIETEN has repeated the density measurements and has corroborated Strelkow's results. From BERRY's work therefore one can conclude that Schottky defects are present to at most 0.1 mole % at the melting point.

2) The second piece of evidence is provided by the measurements of COMPTON at Urbana. COMPTON measured the self-diffusion coefficient of Ag^+ in AgCl and the conductivity of the salt in the intrinsic range below the melting point. He found that the ratio

$$\frac{D_T}{kT\sigma/Ne^2} \quad \text{is} \quad \sim 0.5\,,$$

N is the number of cation sites per unit volume.

This result can be explained following the model proposed by MCCOMBIE and LIDIARD and treated in further detail by COMPAAN and HAVEN. One

supposes that Frenkel defects are formed below the melting point and that the interstitials move by the interstitialcy rather than by the direct mechanism i.e. that the interstitials move by displacing normal ions. In collinear jumps of this sort, the electric charge moves twice the distance the radioactive interstitial atom moves when the latter jumps. If one assumes that all the interstitialcy jumps are collinear, that all the current is carried by interstitials and if one takes into account correlations between successive interstitial jumps, one finds from the Nernst-Einstein relation

$$\frac{\sigma_i}{D_i} = \frac{N_i e^2}{kT} ,$$

where the subscript i refers to interstitials, that

$$\frac{\sigma}{D_T} = 3\frac{Ne^2}{kT} .$$

If we take into account the contribution to the conductivity and to D_I due to vacancy motion, it is found that

$$\frac{\sigma}{D_T} = 2.3\frac{Ne^2}{kT} .$$

The experimental coefficient is about 1.7. The theoretical coefficient may be reduced if one assumes that a certain fraction of interstitialcy jumps are non-collinear.

Some additional experiments of TELTOW might be mentioned at this point. He dissolved Ag_2S in AgCl and AgBr. It was demonstrated that the S^{--} ion dissolves substitutionally with the formation of a Ag interstitial ion: the conductivity increased for increasing concentrations of impurity. Because of the limited solubility of Ag_2S in the salts the experiments were only of semiquantitative nature but they constitute further evidence that the intrinsic defect in Ag halides is the Frenkel type below the melting point.

Some further points may be mentioned:

1) HILSCH and co-workers coevaporated Ag and AgCl at liquid Nitrogen temperatures. The absorption spectrum of such a specimen showed what is presumably an F band at 3 750 Å (an F band is observed in an alkali halide specimen prepared in the same way). When the crystal is heated to room temperature, the band disappears and

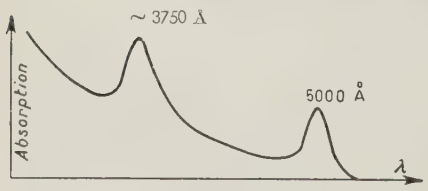


Fig. 18.

is replaced by a colloid band at 5000 Å. The process presumably taking place involves migration of the F-centers to dislocations when they condense to form colloidal particles of silver. Presumably the contribution to the absorption peak of the colloidal spectrum arising from a given center depends on the way the dislocation to which it is attached runs through the crystal.

2) It is observed in photoconductivity measurements that electrons have a remarkably long life time in the Ag halides, of the order of 10^{-5} s; the time in the alkali halides is 10^{-3} times smaller. If one has a specimen whose thickness is of the order of millimeters, one can make electrons go from one electrode to the other and by measuring the transit time, one measures the mobility of an electron. BROWN and VAN HEYNINGEN at Urbana have made such measurements on very pure single crystals. The results will be discussed by Dr. BROWN. in his seminar. At low temperatures the lifetime and mobility drop to values like those in the alkali halides. One must assume that at high temperatures electrons are scattered by lattice vibrations while at low temperatures a new scattering or trapping mechanism sets in. This mechanism is not well understood at present.

3) There has been some investigation of the nature of holes in the Ag halides. It must be mentioned however that all direct measurements of hole conduction reported so far are highly suspect.

Some of the most important experiments investigating the nature of holes in the Ag halides were carried out by LUCKEY and WEST at the Kodak Laboratories in Rochester. They measured the evolution of halogen gas when layers of various thicknesses were irradiated with light of varying wave length. One of the results of this work is that the distance holes are able to migrate is of the order of 1 micron perhaps when bound to positive-ion vacancies.

Holes have also been studied in Ag halides containing a stoichiometric excess of halogen. At Kodak stoichiometric crystals were placed in an atmosphere of the halogen. The specimen then displays electronic conductivity. Presumably a silver atom diffuses out to join a halogen at the surface leaving a hole and a silver-ion vacancy behind. It would be valuable if the Hall effect were investigated in such specimens to prove that the conductivity is of p-type. The Hall effect may be large enough to be measurable.

V.

Metals.

The first attempt to discuss lattice defects in metals quantitatively was made by HUNTINGTON and SEITZ in 1941 and concerned metals with large d-shells (noble metals), specifically copper. More extensive calculations on

the energy required to form vacancies in monovalent metals have recently been carried out by FUMI.

Metals do not display electrolytic conductivity so that the experimental evidence comes to a large extent from diffusion and specialized techniques such as those of spin resonance.

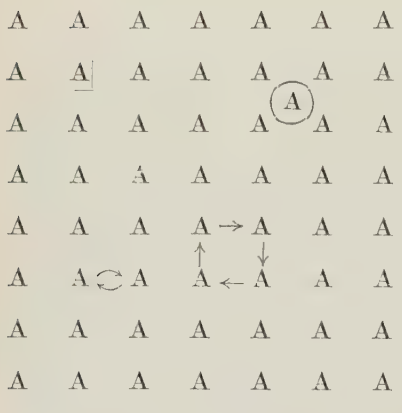
In the one-dimensional case the following diffusion equation is normally valid for macroscopic dimensions:

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left[D(c)\,\frac{\partial c}{\partial x}\right],$$

where $c(x,t)$ is the concentration of solute and $D(c)$ is the diffusion coefficient. This is reduced to a normal differential equation by putting: $\eta = x/\sqrt{t}$,

$$\eta\,\frac{\mathrm{d}c}{\mathrm{d}\eta} = \frac{\mathrm{d}}{\mathrm{d}\eta}\left[D(\eta)\,\frac{\mathrm{d}c}{\mathrm{d}\eta}\right].$$

Diffusion involves lattice imperfections. Three mechanisms may be considered:



It is now generally believed that the vacancy mechanism is the basic one.

Some years ago ZENER reinvestigated the possibility of direct exchange diffusion and pointed out that in copper a ring mechanism, whereby many atoms take place in an exchange process, may be competitive with vacancy diffusion. Also interstitial diffusion may be important in some cases, particularly in radiation-damaged materials.

1. – Diffusion experiments.

Experiments on diffusion in alloys, salts and pure metals became of interest from 1925 onward. The most accurate investigations in metals are

recent and concern self-diffusion in noble metals. We quote some accurate data for Cu, Ag and Au obtained by LAZARUS, SLIFKIN and their co-workers.

Metal	Activation energy
Cu	42 140 cal/mole
Ag	45 500
Au	39 250

The best evidence that diffusion occurs through either vacancy or interstitial mechanism is given by the Kirkendall effect. A block of α-Brass (70% Cu+30% Zn) is surrounded by Cu and markers are placed at the interface. On heating to high temperatures the markers are observed to move inward as the diffusion proceeds. This fact cannot be explained by interchange mechanism and suggests that a defect mechanism must be assumed. Several investigations of the Kirkendall effect were carried out with different couples of metals and alloys and lead to the same conclusions (CORREA DA SILVA and MEHL, ALEXANDER, BALLUFFI and SEIGEL and others).

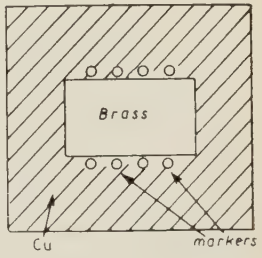


Fig. 19.

DARKEN proposed the following explanation: let us assume that each atom has its own diffusion coefficient and let D_a and D_b be the diffusion coefficients of A and B atoms involved in the Kirkendall effect. We shall also assume that the intrinsic density has the equilibrium value in each small volume in the specimen. The chemical diffusion coefficient is then given by $D_c = f_a D_b + f_b D_a$, where f_a and f_b are concentrations. This quantity should be placed into the basic diffusion equation. The velocity of the marker shift is given by the expression

$$V = \frac{I_a + I_b}{C} = -(D_a - D_c)\frac{\partial C_a}{\partial x}.$$

Darken's result can be derived differently. We assume that there is a current of vacancies I_V from Cu to brass:

$$I_V = -D_V \frac{\mathrm{d}N_V}{\mathrm{d}x} + N\,k\,\frac{\mathrm{d}N_a}{\mathrm{d}x}.$$

The first term is the Fick term and the second arises from the gradient of chemical concentration. Darken's equation may be derived if one assumes that the vacancies are in local equilibrium. If the mechanism is an interchange process, $D_a = D_b$ and no effect would be observed.

Dislocations presumably act as sinks for vacancies maintaining local equilibrium, as well as they can. They work effectively on a macroscopic scale, but may fail on a microscopic one.

The Kirkendall effect can give additional information about defects: in some couples of metals porosity has been noticed near the diffusion boundary. That means that not all vacancies which have crossed the boundary disappear at dislocations: some of them condense to form voids. By studying this effect one can hope to estimate the number of jumps which a vacancy makes before it disappears.

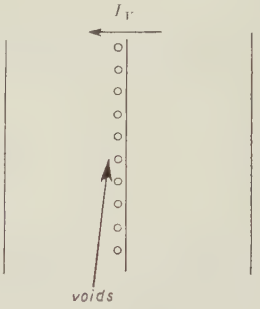


Fig. 20.

2. – Quenching experiments.

We remarked in the first part, that changes in resistivity of gold due to quenched in vacancies were measured by KAUFFMANN and KOEHLER in 1953. Experiments were carried out later by other investigators with several metals.

Results obtained by BAUERLE using brine quenching with gold wires show unusual consistency and clarity. For practical purposes the wires could be considered to be single crystals because the size of grains was of the same order as the diameters of wires. Quenching was performed in the range from 450 °C to 900 °C. It was found that $\log \Delta\varrho$ depended linearly upon $1/T$ and the activation energy ε_f was found to be $(0.96 \pm 2\,\%)$ eV. Using Okkerse's result for the activation energy ε_d for self-diffusion, we find that the migration energy of single vacancies should be

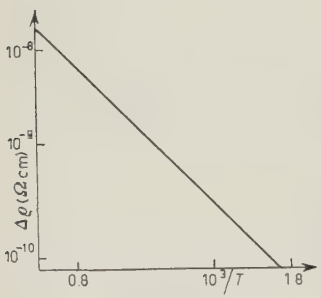


Fig. 21.

$$\varepsilon_m = \varepsilon_d - \varepsilon_f = 1.71 - 0.96 \simeq 0.75 \text{ eV} .$$

Annealing experiments show a linear relationship between the log of residual resistivity and time for quenching temperatures lower than 700 °C. This shows that vacancies probably disappear by jumping into dislocations:

the number ω of jumps they make before dying is $\sim 10^7$ according to annealing data. The small deviation from linearity near $t = 0$ is presumably due to depletion of vacancies near dislocations.

The volume of specimens diminishes during annealing in the same manner as resistivity. It was found that $\Delta\varrho/(\Delta v/v) \simeq 3.3 \cdot 10^{-4}\ \Omega\,\mathrm{cm}$, Therefore one finds that the resistivity change due to 1 % vacancies is about 3.3 $\mu\Omega$cm, if one assumes the lattice does not shrink because of vacancies. Actually the lattice probably shrinks around a vacancy. According to Huntington's calculations this contraction may be important. We may conclude that

$$1.3\ \mu\Omega\mathrm{cm} \leqslant \Delta\varrho \text{ due to } 1\,\% \text{ vac} \leqslant 3.3\ \mu\Omega\mathrm{cm}\,.$$

Other quenching experiments are due to Bradshaw and Pearson (Pt: $E_f = 1.4$ eV, $E_m = 1.1$ eV) and to Panseri, Gatto and Federighi (Al: $E_f = 0.76$ eV). Annealing experiments in Al were performed by De Sorbo and Turnbull ($E_m = 0.37$ eV).

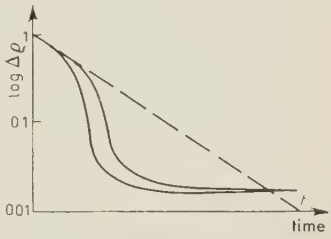


Fig. 22.

It is a remarkable fact that the annealing experiment showed a different mechanism when the specimens were quenched at temperatures above 700 °C. For quenchings at 800 °C or more $\Delta\varrho$ did not anneal completely and curves of $\Delta\varrho$ against t were not exponential. Probably the vacancies generate their own dislocations at higher concentrations. As the figure shows, the resistivity starts to follow the exponential law but then falls more rapidly. It is likely that vacancies evaporate at high temperatures after having been captured at dislocations; consequently their life in the metal is increased. It is possible that ω becomes as large as 10^{10} or 10^{11}. It has been also pointed out by Panseri, Gatto and Federighi that impurity atoms are effective in capturing vacancies in Al.

VI.

Lattice defects in non-polar semiconductors.

1. – In this Section attention is focused on a few topics connected with lattice defects in non-polar semiconductors. Self-diffusion measurements in Ge were carried out by Portnoy, Letaw and Slifkin using a radioactive Ge isotope. Because of the very high activation energy for self-diffusion they had to carry out the measurements near the melting point of Ge, using very thin specimens. The accuracy of the results obtained in such a way is not

as high as can be achieved in the noble metals at present:

$$D_{Ge} = 7.8 \exp\left[-\frac{68\,500}{kT}\right] \text{cm}^2/\text{s} .$$

2. – Let us consider the mechanism of diffusion. The quenching experiments of MAYBURG and LOGAN are interesting in this connection. MAYBURG quenched a Ge crystal (20 mm × 3 mm × 1 mm) from high temperatures by air cooling and observed new acceptor states, which he associated with lattice defects. His quenching rate was 100 °C/s. He derived from his results an imperfection density of

$$\frac{n_a}{N} = 6.7 \exp\left[-\frac{46\,100}{kT}\right],$$

where n_a is the number of defects/cm³, N the number of Ge atoms/cm³. It may be mentioned that these defects change Ge from n- to p-type material. MAYBURG observed a very complicated and incomplete annealing.

LOGAN repeated these experiments. He could reach higher quenching rates in oil, and obtained

$$\frac{n_a}{N} = 29 \exp\left[-\frac{46\,000}{kT}\right].$$

The activation energies derived by both investigators agree although the pre-exponential factors differ. This may be associated with the smaller quenching rates achieved by MAYBURG. Logan's specimen annealed almost completely. Because of this he believed he had measured an effect different from that of MAYBURG. However Mayburg's samples probably were not as pure as Logan's and this would explain the complicated annealing MAYBURG observed. One can estimate the activation energy ε_m of motion and the activation energy ε_f of formation of defects from the activation energy of the self-diffusion ε_d: $\varepsilon_d = 68\,000$, $\varepsilon_f = 46\,000$, $\varepsilon_m = 22\,000$ cal/mole. The experiments suggest that the defects involved are either vacancies or interstitials, probably the former.

3. – The study of the diffusion of copper in Germanium led initially to a confusion of concepts. The diffusion could be very rapid in general but was found to depend upon the density of dislocations by FULLER, DITZENBERGER and WOLFENSTEIN and others. Although it might appear that the copper diffused along dislocation lines, in explanation of the more rapid diffusion in crystals containing relatively high densities of dislocations, quantitative estimates showed that the diffusion rate along the imperfections would have to be impossibly large. The clew to the riddle was provided by TWEET and GAL-

LAGHER who found that one type of migration at a low level of concentration, important for the decoration of dislocations by copper, took place more rapidly the more nearly perfect the crystal.

FRANK and TURNBULL provided the synthetic co-ordinating analysis. They proposed that copper could be present in germanium in two forms, namely substitutionally and interstitially. The copper diffuses very rapidly along interstitial positions if it does not meet lattice defects in the Germanium, presumably vacancies. If it does, the copper atoms fall into the vacancies and are trapped substitutionally for a period of time. The vacancies so occupied will eventually be regenerated by dislocations, the actual time required depending upon the density of dislocations. Thus when the dislocation density is low, the first wave of copper atoms occupy the existing vacancies which are replaced slowly and additional copper diffuses interstitially very rapidly. Eventually the equilibrium ratio of interstitial and substitutional copper is achieved by the generation of vacancies at the small number of dislocations, however the substitutional diffusion of copper is slow. The situation is completely inverted if there are many dislocations. The interstitial diffusion is then impeded by the rapid replacement of vacancies and the equilibrium concentration of substitutional copper is achieved relatively rapidly in the advancing front of copper.

It is evident that the early measurements of diffusion focussed attention on the migration of the substitutional concentration which is about a hundred times larger than the interstitial, at equilibrium. The present estimates of the interstitial and substitutional concentrations and diffusion coefficients are given in the following table:

Interstitial Cu		Substitutional Cu	
D_i	10^{-3} cm²/s	D_s	10^{-5} cm²/s
C_i	10^{-9}	C_s	10^{-7}

It should be noted that the Frank-Turnbull model of the process would be equally valid if the natural disorder in the germanium consisted of interstitial atoms, with which the copper combined, rather than vacancies. Actually it is known that copper normally acts as an acceptor with three vacant levels, as Professor FAN mentioned in his lectures. Hence it appears most natural to believe that the monovalent copper responsible for the acceptor states in germanium is in substitutional positions, where it would naturally accept three electrons to complete its shells as a quadrivalent stand-in for germanium. Presumably an interstitial copper atom acts as a donor.

VII.

Radiation damage in metals.

The study of radiation effects in materials can be divided into two broad periods. Before 1943 investigations were concerned mainly with biological effects. Subsequently, after Wigner's suggestion that materials in reactors could undergo substantial changes in their own properties, many groups started to work in this field. It was soon realized that different solids behave differently; however in this part we shall be mainly concerned with the effects in metals. Most simple metals must be irradiated below room temperature to preserve a major fraction of the defects produced by irradiation.

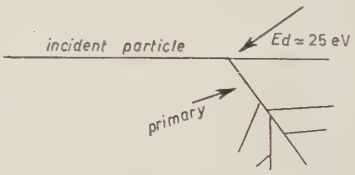


Fig. 23.

Let us consider what happens when a fast charged particle passes through matter. Most of the energy is dissipated in electronic excitation. The residual energy (about 0.1 % of the total energy) is expended in thermal excitation as a result of Coulomb encounters with the nuclei. Some collisions are effective in displacing atoms permanently from their lattice sites. This requires that about 25 eV be transmitted to an atom as translational energy. High temperature regions are also produced in such collisions; interstitials and vacancies will be produced in this process.

The first measurements on metals outside reactors were made with α-particles, deuteron and proton beams; in these cases the average energy transmitted to a knock-on atom which leaves its site is sufficient to make the primary produce secondary displacements. About five secondaries may be expected on the average.

A substantial advance in the field was made when MILLS pointed out that electrons having about 1 MeV should be able to produce displacements. MILLS' proposal proved to be quite convenient for obtaining values of the displacement energy. Actually if the energy of the knock-out atom is not much larger than E_d, the accompanying thermal spikes should be small. Since electron excitation presumably produces no effects in metals, the results can be described entirely in terms of displacements of single atoms.

Fig.24

In the early work, attention was focused mainly on the production of interstitials and vacancies. A very simple model in which the atoms were considered in a square well whose height is about 25 eV was used. The secondary effects could be treated quantitatively in this model. FRENKEL pairs are produced when atoms are knocked out of the well.

It was also realized by BRINKMAN that substantial additional effects may occur in small regions containing a high density of defects. The mean free paths of recoil atoms in heavy materials happen to be of the order of one atomic distance, so that Frenkel pairs can be produced in a small volume. BRINKMAN termed such regions displacement spikes and emphasized that the damaged region may be in a highly chaotic state.

Fission neutrons impart sufficient energy to atoms heavier than Al to make them produce large displacement spikes. Moreover most of the recoil energy is lost by collision. Practically only primary atoms can be produced with electrons near the threshold.

1. – Analysis of results.

After the earliest experimental work on radiation damage it was realized that very low temperatures were needed to preserve the effects in metals because of normal annealing. MARX, COOPER and HENDERSON irradiated copper at liquid nitrogen temperature with 12 MeV deuterons. It is noticed (fig. 25, curves A and B) that the obtainable changes in resistivity depended upon the irradiation temperatures, which means that some annealing took place even near 100 °K. Curve C is corrected for annealing but still bends down. This effect of non-linearity between $\Delta\varrho$ and Φ (independent of thermal annealing) is called radiation annealing.

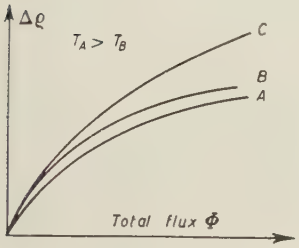


Fig. 25.

The origin of radiation annealing is still not apparent and not well understood. Two possible explanations of it are as follows:

1) a displaced atom may find a vacancy and occupy it;

2) disordered regions may overlap for high Φ, so that the effect of prolonged irradiation is smaller than the initial effect.

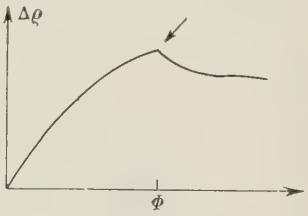


Fig. 26.

Annealing experiments after irradiation at liquid N_2 temperature were carried out by MARX and by OVERHAUSER, who studied the way in which resistivity recovered as a function of temperature.

It was found that annealing takes place almost continuously, indicating that there is a range of activation energies rather than a single activation energy. In Fig. 28 the principal activation energy E is plotted against T: it is apparent that a broad range of activation energies, ranging from 0.2 to

0.6 eV was found. Above -60 °C values of E are practically independent of temperature and are ~ 0.7 eV. This value for the annealing energy has been associated with migration of vacancies, owing to the fact that migration

Fig. 27. – A typical $\Delta\varrho$ versus T curve as a specimen is warmed in stages.

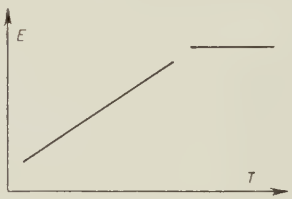


Fig. 28. – Schematic plot of the activation energy effective at each temperature.

of vacancies takes place in the same range of temperature. This conclusion cannot be considered to be proved from these measurements alone, however the results of quenching experiments support the hypothesis.

Overhauser was able to measure the energy released during annealing of Cu samples which were warmed from liquid nitrogen temperature. The ratio between the stored energy U and the decrease in resistivity $\Delta\varrho$ was found to be

$$\frac{U}{\Delta\varrho} = 1.7 \text{ cal/g}\mu\Omega\text{cm}.$$

Let us assume that about 5 eV are released in interstitial + vacancy recombination: then we find from the preceeding ratio that the resistivity associated with 1% displaced atoms should be $\simeq 10\ \mu\Omega$cm. Actually changes in resistivity due to interstitials ($\Delta\varrho_I$) and vacancies ($\Delta\varrho_V$) were calculated by Dexter, Jongenburger and Blatt. They found that

$$\Delta\varrho_I\ (1\%) \simeq \Delta\varrho_V\ (1\%) \simeq 1.5\ \mu\Omega\text{cm}\,.$$

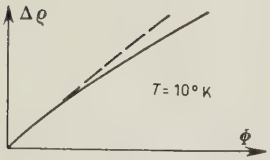


Fig. 29.

Thus the situation was not entirely satisfactory.

Cooper, Koehler and Marx subsequently performed irradiations of Cu, Ag and Au with 12 MeV deuterons at 10 °K. In this case a plot of $\Delta\varrho$ versus Φ is almost linear, depending upon Φ according to the formula $\Delta\varrho = A(1 - e^{-K\Phi})$, where A and K are constants.

In samples irradiated at 10 °K, the annealing mechanism is not the same as in the case in which irradiation was carried out at higher temperatures. There is evidence of an annealing energy of about 0.1 eV. Actually Huntington's

calculations indicate that the migration energy of interstitials is probably less than 0.5 eV; it was therefore assumed tentatively that the experimentally found $E_m \sim 0.1$ eV is associated with migration of interstitials. Some of the interstitials may be expected to persist and anneal together with vacancies at higher temperatures. This interpretation was subject to some doubt. It was proposed that recrystallization of spike was reponsible for the low temperature stage of annealing.

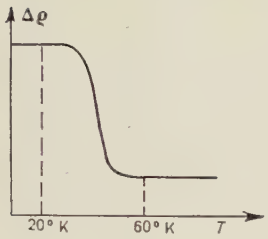


Fig. 30.

In recent years CORBETT, DENNEY, FISKE and WALKER at the General Electric Laboratories irradiated highly pure copper with 1.35 MeV electrons. In this case the secondary displacements should be almost negligible and the effects of displacement spikes should be nearly absent. The annealing is extremely simple. Presumably the displaced atoms do not go far from their lattice sites and direct pair annealing predominates.

Recently PALMER and MAGNUSON performed irradiations and annealing measurements on deu teron irradiated specimens. Several annealing peaks were evident the most prominent being near to 0.1 eV. The annealing reaction was presumably monomolecular, with a frequency coefficient $\nu_0 \simeq 10^{11}\ s^{-1}$, which is less than the atomic frequency. This shows that the migration of defects occurs over many atomic distances and that the stage of annealing involving activation energies near 0.1 eV may be connected with annihilation of distant pairs.

Very significant results were obtained by measuring the changes in volume (ΔV) and in lattice parameter of copper during irradiation and annealing. BALLUFFI, SIMMONS and VOOK found that within experimental accuracy the relative change of lattice parameter is equal to that of linear dimensions both during irradiation and during annealing. This is expected for Frenkel and not for Schottky disorder. Thus the measurements support the view that:

1) Frenkel defects are indeed produced by bombardment,.

2) Pair annihilation takes place after diffusion of one defect.

From the ratio $\Delta\varrho/\Delta V$ one can determine $\Delta\varrho_V$ and $\Delta\varrho_A$. We can make a compromise between these and previously discussed results and make the reasonable assumption that $\Delta\varrho_V$ and $\Delta\varrho_A$ are quite closely equal:

$$\Delta\varrho_A \simeq \Delta\varrho_V \simeq 3.0\ \mu\Omega\text{cm} \qquad \text{for } 1\,\% \text{ defects}$$

and that about 3 eV are released in an interstitial-vacancy recombination.

One very important experimental result throws some doubt on these con-

clusions. BLEWITT and his colleagues performed fast neutron irradiation on copper in the Oak Ridge reactor and measured the stored energy release on annealing at low temperatures. They obtained a negative result and suggest that the value may be less than 1.7 cal/g/$\gamma\Omega$cm. If we lower the value of $U\,\Delta\varrho$ correspondingly the consistent picture discussed above disappears. The stored energy observed at Oak Ridge seems smaller than one would expect from interstitial-vacancy annihilation. Conversely, if one accepts the result, the resistivity associated with 1% of displaced pairs must be assumed very large compared to the values discussed above. This contradiction awaits resolution before we can feel confident of our understanding of radiation effects in the simple metals.

BIBLIOGRAPHY

I.

Imperfections in Nearly Perfect Crystals (New York, 1951).
M. BALKANSKI and I. BROSER: *Report of Garmisch Meeting*, 1956.
G. DIEMER and W. HOOGENSTRAATEN: *Journ. Phys. Chem. Solids*, **2**, 119 (1957).
J. A. BAUERLE and J. S. KOEHLER: *Bull. Am. Phys. Soc.*, **1**, 145 (1957).

II.

A. B. LIDIARD: *Handbuch der Physik*, **20**, 256 (1956).
F. SEITZ: *Rev. Mod. Phys.*, **26**, 7 (1954).
N. F. MOTT and R. W. GURNEY: *Electronic Processes in Ionic Crystals* (Oxford, 1949).
Y. HAVEN: *Report of* 1954 *Bristol Conference* (*Proceedings of Physical Society of London*, 1955).

III.

Same as II.
S. AMELINCKX, W. BONTINCK, W. DEKEYSER and F. SEITZ: *Phil. Mag.*, **2**, 355 (1957).

IV.

F. SEITZ: *Rev. Mod. Phys.*, **23**, 328 (1951); *Photographic Sensitivity* (Tokyo, 1956).
Forthcoming article by F. BROWN and F. SEITZ in 1957 *Conference on Photographic Sensitivity in Japan*.

V.

F. SEITZ: *Journ. Phys. Soc. Japan*, **10**, 679 (1955).
R. W. BALLUFFI and L. L. SEIGLE: *Journ. Appl. Phys.*, **25**, 607 (1954).
R. RESNICK and L. SEIGLE: *Journ. Met.*, **9**, 87 (1957).

VI.

H. Y. Fan: *Solid State Physics*, Vol. **1**.
W. M, Portnoy, H. Letaw and L. Slifkin: *Phys. Rev.*, **93**, 892 (1954); **102**, 636 (1956),
S. Mayburg: *Phys. Rev.*, **95**, 38 (1954); **98**, 1134 (1955).
R. A. Logan: *Phys. Rev.*, **101**, 1455 (1956).
C. S, Fuller, and J, A. Ditzenberger: *Journ. App. Phys.*, **28**, 40 (1956).
A. G. Tweet and C. J. Gallagher: *Phys. Rev.*, **103**, 828 (1956).
C. Frank and D. Turnbull: *Phys. Rev.*, **104**, 617 (1956).
F. Van der Maasen and J. A. Brinkman: *Journ. Electrochem. Soc.*, **102**, 229 (1955).

VII.

F. Seitz and J. S. Koehler: *Radiation Effects in Solids*, in *Solid State Physics*, Vol. **2** (1956).
J. W. Corbett, J. M. Denney, M. D. Fiske and R. M. Walker: *Phys. Rev.*, **104**, 851 (1956).
R. O. Simmons and R. W. Balluffi: *Bull. Am. Phys. Soc.*, **112**, 151 (1957).
C. A. Klabunde, T. H. Blewitt, R. R. Coltman and T. S. Noggle: *Bull. Am. Phys. Soc.*, **2**, 150 (1957).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Effets des imperfections cristallines sur la diffraction des rayons X.

A. GUINIER

Conservatoire National des Arts et Métiers - Paris

1.

Théorie générale.

Il est bien connu et évident que l'application la plus importante de la diffraction des rayons X a été l'étude de la structure des cristaux: il s'agit de trouver comment les atomes sont disposés dans la *maille* que l'on suppose être répétée, identique à elle-même, par toutes les translations du réseau cristallin périodique. Pour le physicien du solide, la structure du cristal idéal est certes nécessaire à connaître, mais elle est aussi insuffisante. Savoir que l'aluminium cristallise dans le système cubique à faces centrées avec une maille de 4.04 Å est bien peu de chose pour celui qui cherche à comprendre les propriétés physiques de ce métal. Heureusement, les méthodes classiques des rayons X apportent d'autres données, par exemple sur l'état de cristallisation, la texture cristalline, les légères variations de paramètre sous l'influence des tensions, etc.

Mais c'est un autre domaine d'application des rayons X qui sera le thème de ces conférences. L'on sait que les défauts du cristal jouent un rôle de premier ordre dans la physique du métal. Nous allons étudier quelles sont les données des rayons X sur ces défauts. Disons tout de suite que l'on est loin de pouvoir obtenir une image complète des structures irrégulières. Mais, étant donné notre pauvreté en moyens d'observation directe dans ce domaine, des renseignements même incomplets ou peu précis sont néanmoins intéressants. C'est ce qui explique que de très nombreux travaux tant théoriques qu'expérimentaux ont été récemment publiés sur ce sujet.

Il ne saurait être question ici d'entrer dans les détails de la théorie de la diffraction des rayons X par des cristaux imparfaits, car les calculs devraient être si résumés qu'ils deviendraient difficiles à suivre. Notre propos sera donc seulement de préciser les possibilités des rayons X, en vue de faire connaître aux physiciens du solide les problèmes qu'ils peuvent poser aux techniciens

des rayons X et la nature des réponses qu'ils peuvent espérer. On trouvera dans les références bibliographiques la justification des résultats que nous nous contenterons d'énoncer. En particulier, pour les rappels des notions élémentaires et pour les développements spécialisés, nous renverrons le lecteur à notre traité récemment paru (GUINIER [1] qui sera désigné per *Radiocrist.*).

1. – Cristal parfait et cristal mosaïque.

Soit un cristal parfait de très petite taille (de l'ordre du micron). La théorie cinématique de la diffraction fait usage de la notion d'espace réciproque (*Radiocrist.*, p. 101). A partir du centre de cet espace, nous traçons le *réseau réciproque* du réseau cristallin donné, dont voici la propriété fondamentale: un nœud quelconque R_{hkl} correspond à la famille des plans réticulaires (hkl) de réseau cristallin. Le vecteur $\boldsymbol{OR}_{hkl}$ joignant l'origine au nœud est normal à ces plans et $OR_{hkl} = 1/d_{hkl}$. Une expérience de diffraction en rayonnement monochromatique est caractérisé par un vecteur $\boldsymbol{s} = (\boldsymbol{S}/\lambda) - (\boldsymbol{S}_0/\lambda)$, où λ est la longueur d'onde, $\boldsymbol{S}_0$ et $\boldsymbol{S}$ les vecteurs unitaires suivant les directions du rayon incident et de la direction d'observation (Fig. 1).

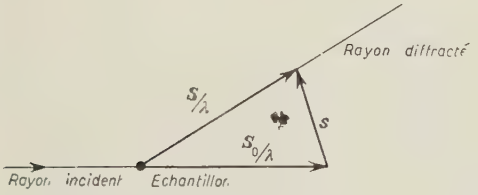


Fig. 1. – Définition du vecteur $\boldsymbol{s}$ à partir des directions des rayons incident et diffracté.

Les phénomènes de diffraction sont décrits simplement par la loi d'Ewald. Il y a un rayon diffracté dans la direction $\boldsymbol{S}$, si l'extremité du vecteur $\boldsymbol{s}$, dont l'origine est placée au centre du reseau réciproque, coïncide avec un nœud du réseau réciproque (Fig. 2). L'intensité diffractée est proportionnelle au carré du facteur de structure F_{hkl} attaché à ce nœud. F_{hkl} dépend de la position des atomes dans la maille. Quand l'extrémité de $\boldsymbol{s}$ est en dehors de ces nœuds, l'intensité diffractée est rigoureusement nulle. En réalité ce que nous appelons nœud est un très petit domaine entourant le nœud géométrique, ce domaine étant d'autant plus étendu que le cristal est plus petit.

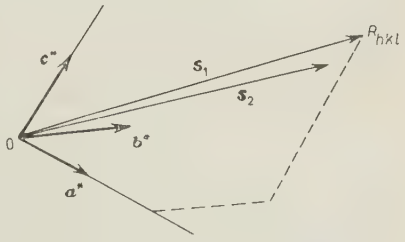


Fig. 2. – Détermination des directions des rayons diffractés par un cristal parfait par la construction d'Ewald. Il y a diffraction dans le cas 1 (vecteur $\boldsymbol{s}_1$ au noeud R_{hkl}) et non dans le cas 2. $\boldsymbol{a}^*$, $\boldsymbol{b}^*$, $\boldsymbol{c}^*$ sont les axes du réseau réciproque.

En fait, un échantillon diffractant est toujours de dimension bien supérieure au micron. Si le cristal dans toute son étendue est rigoureusement parfait, les phénomènes sont compliqués par les interférences entre rayons incidents et diffractés qui ne sont négligeables que si le nombre de plans diffractants est assez petit (< 1000). Pour le gros cristal parfait, suivant la théorie dynamique (James [2]), les conclusions de la théorie élémentaire pour la localisation du faisceau diffracté sont encore valables mais l'intensité est profondément modifiée. L'expérience montre que les conclusions de cette théroie ne sont pas vérifiées sauf avec quelques cristaux exceptionnels (calcite, diamant). Ceci ne peut s'expliquer que par la présence dans presque tous les cristaux d'imperfections qui rompent la cohérence des plans réticulaires à des distances de l'ordre du micron.

C'est là le premier résultat des rayons X sur les imperfections cristallines. Darwin avait imaginé le modèle de la *mosaïque* composée de petits blocs très légèrement décalés ou désorientés de façon à ce qu'ils diffractent de façon incohérente. Aujourd'hui, nous pouvons assimiler les blocs mosaïques hypothétiques aux blocs limités par les réseaux des lignes de dislocations que de nombreuses expériences ont révélées (réseau de Frank). Les dimensions des blocs sont bien de l'ordre de grandeur voulue. Mais il n'est pas certain que l'on puisse admettre, d'une part, qu'il n'y a pas d'interférence entre les ondes diffractées pour deux blocs contigus et, d'autre part, qu'un bloc élémentaire est assez petit pour que la théorie cinématique s'applique sans correction.

De fait, les mesures du pouvoir réflecteur des cristaux réels ont donnée des nombres compris entre les deux valeurs limites du cristal parfait (théorie dynamique) et du cristal mosaïque (théorie cinématique), valeurs qui sont séparées l'une de l'autre par un facteur au moins de l'ordre de 10. Les effets de l'existence des lignes de dislocation dans le « bon cristal » doit donc être facilement perceptible par des mesures du pouvoir réflecteur. Malheureusement la théorie n'a pas encore été faite: aussi la méthode n'a pas encore été utilisée. Elle pourrait être utile pour l'étude de cristaux presque parfaits.

Dans la plupart des cas, on peut considérer le cristal comme parfaitement mosaïque. C'est ce que nous admettons dans la suite: nous ne nous occuperons plus de ces défauts à grande échelle, mais des irrégularités à l'échelle atomique réparties dans tout le volume du cristal.

2. – Défauts de périodicité d'un cristal.

On peut décrire de deux façons les faibles perturbations de périodicité à partir d'un cristal parfait. Pour la simplicité de l'écriture, nous considérerons un cristal avec un seul atome par maille, mais la généralisation à des mailles complexes ne présente pas de difficultés. Les perturbations peuvent venir.

soit de la substitution d'un autre atome à l'atome du cristal donné, soit du déplacement de l'atome hors du nœud, soit de l'ensemble des deux causes.

2·1. *Perturbations du 1er genre.* – Le cristal réel est défini à partir d'un *cristal parfait moyen*. La position du n-ième atome est donnée par $\boldsymbol{x}_n + \boldsymbol{\Delta}_n$, $\boldsymbol{x}_n$ étant le vecteur du réseau moyen de coordonnées n (n désignant le triplet des 3 entiers n_1, n_2, n_3) et le déplacement $\boldsymbol{\Delta}_n$ étant petit devant les dimensions de la maille, quel que soit n. Le réseau moyen est choisi de façon que $\sum \boldsymbol{\Delta}_n = 0$, la somme étant étendue à tous les atomes du cristal. Dans le cristal réel, les atomes peuvent varier de nature, de maille à maille; le facteur de diffusion est fonction de la position de la maille, F_n. Dans le cristal moyen, toutes les mailles sont occupées par un même atome fictif dont le facteur de diffusion est la moyenne des facteurs F_n.

Comme exemples, citons pour les déplacements purs, le désordre dû à l'agitation thermique, et, pour les substitutions pures, la solution solide non parfaitement ordonnée.

2·2. *Perturbations du 2ème genre.* – Un cristal est parfait parce que la disposition des voisins immédiats de tous les atomes homologues est rigoureusement identique. Supposons que cet arrangement local puisse présenter de légères fluctuations par rapport au schéma idéal. En construisant ainsi un cristal de proche en proche à partir d'un centre, on trouve que les fluctuations des positions réelles par rapport à celles d'un réseau régulier croissent avec la distance du centre. *L'ordre à petite distance* est à peu près conservé, mais l'ordre à grande distance a disparu, tandis que dans les perturbations du 1er genre, il n'était que légèrement perturbé.

Nous citerons comme exemple de ce type de désordre les hauts polymères (cellulose): les forces entre chaînes voisines étant assez faibles, on conçoit que leur arrangement mutuel puisse subir des fluctuations qui n'existent pas dans les cristaux minéraux où les forces interatomiques sont bien plus considérables.

3. – Effets des défauts de périodicité sur la diffraction des rayons X.

La formule fondamentale de la théorie de la diffraction est l'expression

$$A(\boldsymbol{s}) = \int \varrho(\boldsymbol{x}) \exp\left[-2\pi i \boldsymbol{s}\cdot\boldsymbol{x}\right] \mathrm{d}v_x \,. \tag{1}$$

Elle exprime l'amplitude diffractée au point de l'espace réciproque défini par le vecteur $\boldsymbol{s} = (\boldsymbol{S}/\lambda) - (\boldsymbol{S}_0/\lambda)$. $\varrho(\boldsymbol{x})$ est la densité électronique dans l'objet diffractant. Si celui-ci est un cristal parfait, $\varrho(\boldsymbol{x})$ est périodique et $A(\boldsymbol{s})$ est,

d'après (1), nul pour toutes valeurs de $\boldsymbol{s}$, sauf si $\boldsymbol{s}$ a pour extrémité un nœud du réseau réciproque. C'est la loi élémentaire que nous avions rappelée à la Sect. **1**.

Si $\varrho(\boldsymbol{x})$ n'est pas périodique, comme dans les cristaux imparfaits, $A(\boldsymbol{s})$ peut avoir une valeur non nulle pour tout point de l'espace réciproque. Néanmoins, si les défauts de périodicité sont faibles, $A(\boldsymbol{s})$ garde une valeur considérable aux nœuds du réseau moyen dont dérive le cristal réel.

Ainsi l'effet des irrégularités d'un réseau est de donner naissance à un rayonnement diffracté quand les conditions de Bragg ne son pas réalisées. Comme ce rayonnement ne se produit pas dans des directions étroitement déterminées, on dit qu'il y a *diffusion* et, considérant l'espace réciproque, on dit qu'il y a *diffusion pour des points situés en dehors des nœuds du réseau réciproque*. Si cette diffusion est concentrée dans un étroit domaine entourant ces nœuds, elle se traduit expérimentalement par un élargissement des taches ou raies de diffraction cristalline; si, au contraire, l'intensité de la diffusion est répartie dans toute la maille, cela se traduit par l'apparition sur les diagrammes de taches très floues ou même d'un voile général se superposant aux taches de diffractions cristallines.

Un théorème général montre que l'intensité de cette diffusion doit se soustraire de l'intensité de la diffraction sur les nœuds du réseau réciproque du cristal non perturbé (*Radiocrist.*, p. 410). Il s'ensuit que les imperfections du cristal ont pour effet de diminuer l'intensité des réflexions cristallines.

Nous pouvons ainsi prévoir de façon simple les trois effets possibles des imperfections: *diffusion continue*, *élargissement des taches de réflexion*, *diminution de l'intensité de réflexion*. Il est important de noter dès maintenant que, même si le nombre des mailles perturbées est relativement grand, les effets précédents sont assez faibles, ce qui rend leur détection malaisée.

Le problème est d'utiliser ces effets pour déterminer la nature de la perturbation du cristal. La difficulté majeure est celle que l'on rencontre dans les recherches de structure en cristallographie: $\varrho(\boldsymbol{x})$ peut se déterminer mathématiquement à partir de l'équation (1) par une transformation de Fourier. Mais l'expérience ne donne pas $A(\boldsymbol{s})$, mais seulement l'intensité, c'est-à-dire $I(\boldsymbol{s}) = |A(\boldsymbol{s})|^2$. On connaît le module de $A(\boldsymbol{s})$ mais on ignore son argument. Physiquement, cela veut dire qu'on n'a aucun moyen de comparer les phases des ondes diffractées dans différentes directions, on ne peut qu'en comparer les intensités.

Rappelons qu'en cristallographie, la transformée de Fourier de $I(\boldsymbol{s})$ donne la fonction de Patterson (*Radiocrist.*, p. 150). Nous allons indiquer ce qui, pour les réseaux perturbés correspond à la fonction de Patterson, dans le cas des imperfections du 1er genre, c'est-à-dire définies dans tout le cristal réel à partir d'un réseau moyen.

Dans la n-ième maille, l'atome est déplacé de $\boldsymbol{\Delta n}$ et a un facteur de dif-

fusion f_n. Nous considérons le cristal réel comme un cristal dont le réseau est le réseau moyen rigoureux mais dont le n-ième nœud est occupé par l'atome fictif de facteur de diffusion $F_n = f_n \exp[-2\pi i \boldsymbol{s}\cdot\boldsymbol{\Delta x}_n]$.

Considérons maintenant un vecteur $\boldsymbol{x}_m$ du réseau moyen et tous les couples des mailles d'indices n et $n+m$ séparées par le vecteur $\boldsymbol{x}_m$. On forme le produit $F_n \cdot F^*_{n+m}$. Soit y_m la valeur moyenne de ce produit dans le cristal. C'est ce paramètre y_m qui est lié à l'intensité mesurée. On démontre la relation (*Radiocrist.*, p. 492)

$$I(\boldsymbol{s}) = \sum_{m=0}^{m=\infty} y_m \exp[2\pi i \boldsymbol{s}\cdot\boldsymbol{x}_m)\,, \tag{2}$$

relation valable sous réserve que le cristal diffractant ne soit pas trop petit.

Cela signifie que si y_m est indépendant de s, l'intensité diffusée observée $I(\boldsymbol{s})$ est la transformée dans l'espace de Fourier d'un espace objet constitué par les paramètres y_m attachés à chaque nœud d'indice m (m_1, m_2, m_3), la valeur de y_m étant fonction de ces indices.

Si les défauts du cristal sont nuls, les mailles sont identiques $(F_n = F)$, donc $y_m = F_n \cdot F^*_{n+m} = F^2$. Dans ces conditions, $I(\boldsymbol{s})$ est nul sauf aux nœuds du réseau réciproque, comme on doit s'y attendre puisque le réseau est régulier.

Dans le cas du réseau perturbé, faisons intervenir la valeur moyenne F de la maille moyenne on peut poser

$$y_m = |\overline{F}|^2 + \Phi_m\,. \tag{3}$$

Il s'ensuit que l'on partage $I(\boldsymbol{s})$ en deux parts: la première est la figure de diffraction du réseau moyen idéal, c'est-à-dire qu'elle ne comprend que les nœuds du réseau réciproque; la seconde est l'intensitè diffusée, I_d telle que

$$I_d(\boldsymbol{s}) = \sum \Phi_m \exp[2\pi i \boldsymbol{s}\cdot\boldsymbol{x}_m]\,. \tag{4}$$

3·1. *Conséquences générales.* – C'est la formule (4) qui résume la théorie de l'effet des imperfections sur les rayons X; elle a été présentée sous de nombreuses formes suivant les cas particuliers envisagés mais ces formes sont équivalentes.

Nous en tirons les conséquences suivantes:

1) L'intensité diffusée dépend de Φ_m; c'est-à-dire non seulement des désordres dans chaque maille, mais de la *corrélation existant entre désordres des mailles voisines.* Un cas extrême est celui où il n'y a aucune corrélation, c'est-à-dire que F_{n+m} est complètement indépendent de F_n. Dans ces con-

ditions

$$\overline{F_n F^*_{n+m}} = \overline{F}_n \cdot \overline{F}^*_{n+m} = |\overline{F}|^2$$

et $\Phi_m = 0$ quel que soit m. Seul le terme Φ_0 est différent de 0. En effet $\overline{F_n \cdot F^*_n} = \overline{|F|^2}$. Donc, d'après (3)

$$\Phi_0 = \overline{|F|^2} - |\overline{F}|^2 .$$

A ce seul terme Φ_0 correspond une intensité de diffusion constante quel que soit s (*). C'est la diffusion, dite de *Laue*, car LAUE en avait fait le calcul dans le cas particulier des solutions solides complètement désordonnées.

Si, au contraire, il y a très forte corrélation entre mailles voisines et que cette corrélation décroît assez lentement avec m, la diffusion $I_d(\boldsymbol{s})$, d'après les propriétés des transformées de Fourier, se trouve concentrée autour des nœuds du réseau moyen.

D'une façon plus générale, on peut décomposer Φ_m dans l'espace du cristal en ondes planes sinusoïdales de vecteur d'onde $\boldsymbol{K}$. On montre qu'en un point de l'espace réciproque, seul intervient l'onde dont le vecteur $\boldsymbol{K}$ est égal au vecteur joignant ce point au nœud du réseau réciproque le plus proche. L'intensité de la diffusion en ce point sera proportionnelle à l'amplitude de la composante de vecteur d'onde $\boldsymbol{K}$ dans le spectre de Φ_m. Donc les courtes longueurs d'onde du spectre auxquelles correspondent de grands vecteur $\boldsymbol{K}$ ($K = 1/\lambda$) sont la cause de la diffusion loin des nœuds et, au contraire, les grandes longueurs d'onde produisent des diffusions ramassées autour des nœuds.

Il est possible mathématiquement de déterminer Φ_m pour toutes valeurs de m à partir de $I_d(\boldsymbol{s})$. C'est le renseignement le plus complet que l'on puisse obtenir sur le désordre du réseau sans aucune hypothèse *à priori*. Mais c'est rarement suffisant pour être exploité de façon intéressante.

3·2. *Désordres planaires et linéaires.* – Cependant des résultats généraux très importants découlent directement de la théorie précédente, en particulier quand la corrélation des désordres est essentiellement anisotrope dans le cristal.

Le désordre est dit *planaire*, quand les plans d'une certaine famille sont conservés parfaitement périodiques et restent parallèles entre eux (*Radiocrist.*, p. 506). Mais ces plans ne se reproduisent pas régulièrement, soit à cause de déplacements d'ensemble du plan, soit parce que les plans successifs sont chargés d'atomes différents.

Ces hypothèses entraînent pour la répartition de la diffusion des conditions

(*) A part une lente décroissance avec s, dû à la décroissance des facteurs de diffusion atomiques.

très précises: la diffusion est nulle, sauf sur les rangées du réseau réciproque normales aux plans réticulaires en arrangement désordonné. C'est un fait qui a été observé dans de nombreux cas. Cela se traduit sur les diagrammes du cristal tournant ou de Laue par des traînées (Streaks), qui ne seraient pas explicables par les lois normales de la diffraction des rayons X. Signalons qu'il ne faut pas confondre ces traînées avec celles qui apparaissent sur un diagramme de cristal tournant quand le rayonnement n'est pas strictement monochromatisé: ces dernières sont des phénomènes normaux de diffraction dus au spectre continu; ce spectre continu complique beaucoup toute étude du désordre cristallin: aussi est-il indispensable dans ce genre d'étude de ne pas se contenter de rayonnement filtré, mais d'utiliser un monochromateur cristallin (Geissler et Hill [3]).

On a aussi observé des désordres *linéaires*, c'est-à-dire des cristaux où des rangées d'une famille restent inaltérées, alors que l'arrangement mutuel des rangées cesse d'être régulier. Dans ce cas, la diffusion est localisée dans les plans réticulaires du réseau réciproque normaux aux rangées en arrangement désordonné.

3·3. *Désordre de substitution et de déplacement.* – Une autre propriété générale de la diffusion des cristaux désordonnés est la suivante: on peut facilement distinguer le désordre du déplacement pur du désordre de substitution pure.

Quand la position des atomes est inchangée et que seule leur nature est modifiée de façon irrégulière, on trouve que les facteurs de désordre Φ_m sont indépendants du vecteur $\boldsymbol{s}$. Donc $I_d(\boldsymbol{s})$, d'après (4), est périodique. La diffusion se reproduit identiquement de maille en maille du réseau réciproque (*).

Au contraire quand on déplace les atomes sans en changer la nature, l'intensité de la diffusion croît de maille en maille à mesure qu'on s'éloigne du centre du réseau réciproque. On peut dire, de façon approchée, qu'en des points homologues dans la maille, $I_d(\boldsymbol{s})$ est proportionnel au carré de la distance s au centre de l'espace réciproque. Physiquement, cela signifie que la diffusion est nulle aux très faibles angles et croît avec l'angle de diffusion.

3·4. *Cas des structures centrosymétriques.* – Ces remarques permettent de déduire immédiatement des conditions simples sur la nature des désordres indépendamment de toute interprétation détaillée. Si l'on fait des hypothèses particulières sur la nature du désordre du cristal, on peut préciser les résultats généraux qui viennent d'être énoncés.

Une des possibilités est de supposer que la zône désordonnée a un centre de symétrie. Dans ces conditions la phase de l'amplitude diffractée, quel que soit $\boldsymbol{s}$, est toujours 0 ou π. L'intensité mesurée donne donc l'amplitude au

(*) Si on la corrige de la lente décroissance des facteurs de diffusion atomique.

signe près. Par raison de continuité, ce signe ne peut changer que si l'intensité s'annule. Il est donc le même dans la tache de diffusion entourant un nœud par exemple (Yin-Yuan Li, Smoluchowski [14]). La connaissance de l'amplitude conduit, par inversion de Fourier, à la densité électronique dans le cristal, donc directement à la répartition des atomes. Mais celle-ci repose sur l'hypothèse de la centrosymétrie: ce qu'on peut dire, c'est que la structure trouvée est celle de la zône perturbé centrosymétrique qui donne des phénomènes de diffraction semblables à ceux qui sont observés; mais cela ne veut pas dire que la structure trouvée soit surement correcte: d'autres structures non centrosymétriques sont possibles pour expliquer les mêmes diagrammes de rayons X.

4. – Les effets de diffraction prévus pour différents types de défauts.

4·1. *Défauts ponctuels*: *lacunes*. – Si on néglige les distorsions autour de la lacune, le cristal avec lacunes peut être considéré comme une solution solide dont le réseau est parfait et aux nœuds duquel se trouvent sans aucun ordre les atomes normaux (facteur de diffusion f), et la lacune (facteur de diffusion nul). La diffusion est une diffusion continue de Laue. Son intensité est très faible étant donné l'ordre de grandeur des concentrations possibles des lacunes, c.

Le pouvoir diffusant (*) du cristal est par atome cf^2: par exemple, pour le cuivre, $f = 28$ donc $I_a = 780c$, quantité qui est de l'ordre de 1 pour $c = 10^{-3}$.

C'est la limite des possibilités expérimentales acutelles. L'expérience doit être faite de façon à minimiser les causes possibles de diffusion; il faut opérer à basse température pour diminuer l'effet de l'agitation thermique et aux faibles angles parce que c'est dans cette région de l'espace réciproque que l'effet Compton est minimal, ainsi que l'effet de l'agitation thermique. A notre connaissance, la détection des lacunes par les rayons X n'a pas encore été tentée. L'expérience serait certainement difficile mais pourrait peut-être réussir dans des cas très favorables.

4·2. *Atomes en substitution*. – Le contraste des nombres atomiques donne une diffusion continue de Laue, détectable dans les meilleures conditions aux faibles angles. Il s'y superpose éventuellement une diffusion due à la distorsion provoquée par la différence de taille des atomes (effet Huang) (Huang [4], Warren, Averbach, Roberts [5], Borie [6]). Elle se produit au voisinage des nœuds du réseau; elle est caractérisée par deux maximums situés sur la droite joignant l'origine au nœud. Cet effet, très faible aussi, a été observé, semble-t-il, par Borie dans les alliages $AuCu_3$.

(*) Le pouvoir diffusant est égal au nombre d'électrons qui, diffusant indépendamment, produirait la même énergie diffusée que l'échantillon dans les mêmes conditions.

4·3. *Atomes interstitiels.* – Il y a une diffusion entre des nœuds qui a été calculée dans le cas des réseaux cubiques à faces centrées pour les différentes positions possibles de l'interstitiel, sa répartition ressemble à l'effet d'un atome de substitution pour certaines positions, mais pour l'une d'elles, assez invraisemblable d'ailleurs, on trouve une tache de diffusion avec maximum au nœud (Kunkazi [7]). Il n'y a pas encore de confirmation expérimentale de ces prévisions.

4·4. *Dislocations.* – Le calcul a été fait pour un cristal en forme de cylindre très long et contenant une dislocation vis ou coin le long de son axe (Wilson [8], Suzuki [9]). Pour la dislocation coin, le calcul très complexe est considérablement aidé par une analogie optique (Willis [10]). En effet le problème est en réalité à deux dimensions, puisque les plans normaux à l'axe sont tous identiques: on fait le modèle du plan d'atomes, sa figure de diffraction optique est identique à celle des rayons X pour un même rapport de la maille du réseau à la longueur d'onde.

Les résultats sunt les suivants:

1) Dislocation coin. Tous les phénomènes de diffraction ou de diffusion sont contenus dans les plans du réseau réciproque normaux à la ligne de dislocation, puisqu'il s'agit d'un désordre typiquement linéaire. Les nœuds sont élargis en taches de forme compliquée (petits anneaux ou forme de 8) dont la dimension est inversement proportionnelle au diamètre du cristal et croissant avec les indices de la tache. Autour du centre, il y a une faible diffusion, concentrée dans le plan normal à la ligne de dislocation passant par l'origine; elle est composée de deux taches floues de chaque côté du centre. Enfin il devrait se produire des traînées diffuses joignant les nœuds du réseau non perturbé parallèlement au plan de la dislocation.

2) Dislocation vis. Le calcul de Wilson prévoit un anneau autour des nœuds, d'autant plus intense que l'ordre est plus élevé. Il n'y a pas de diffusion autour de l'origine. La diffusion loin des nœuds ne peut pas être calculée.

Comment peut-on utiliser ces résultats théoriques? Une chose est assez gênante, c'est que le résultat trouvé dépend du diamètre du cristal. Si celui-ci est de très grande taille et contient de nombreuses dislocations parallèles, Wilson propose d'utiliser le calcul en remplaçant le diamètre du cristal par la distance moyenne entre dislocations, mais ce n'est pas aisément justifiable. Pour ce qui est de la diffusion sur les rangées, l'agitation thermique donne déjà des traînées joignant les nœuds voisins, il est probable que cette diffusion masque complètement celle qui est due aux dislocations.

Reste la diffusion aux petits angles. Il est connu que les métaux écrouis produisent une diffusion autour du centre, fonction de l'écrouissage. L'inter-

prétation de cet effet n'est pas encore bien établie. Blin [11] l'avait attribué à des microcavités, mais ce point de vue a été critiqué (Beeman [15], Lowde [16]). On pourrait aussi chercher à l'attribuer à des dislocations. La distinction serait aisée si les dislocations étaient orientées, car elles devraient donner une diffusion très anisotrope.

Signalons les observations de Franks et Holden [12] sur des monocristaux de cuivre soumis à la fatigue. On observe des traînées étroites de diffusion issues de l'origine le long des axes ⟨111⟩. Là encore l'explication est encore assez ambiguë: des groupes de dislocations doivent intervenir.

En résumé, il est certain que les dislocations donnent des effets assez difficiles à mettre en évidence. Jusqu'ici les expériences n'ont pas été décisives. C'est seulement sur des monocristaux, dans des cas favorables et avec des techniques très soignées qu'on peut espérer tirer des résultats intéressants des méthodes des rayons X.

Les défauts que nous avons examinés jusqu'ici n'intéressent qu'un petit nombre d'atomes: la conclusion à laquelle nous sommes arrivés est qu'ils donnent des effets atteignant à peine la limite des possibilités expérimentales. Heureusement *les rayons X sont bien plus utilisables quand on considère des défauts de dimensions un peu plus grandes.* En effet considérons par exemple des cavités sphériques dispersées dans un milieu homogène, elles produisent une diffusion aux petits angles qui, *pour un même volume total des cavités*, a une intensité maximale proportionnelle au *cube* du rayon (*Radiocrist.*, p. 650). Quand les zônes perturbées atteignent des dimensions de l'ordre, disons d'une dizaine ou centaine d'angström, les rayons X deviennent un instrument sensible de détection, très utile parce qu'on est encore bien au-dessous du domaine de la microscopie optique et que le microscope électronique peut être d'un usage difficile. Nous étudierons en détail, dans les conférences suivantes, les effets de quelques défauts submicroscopiques. Pour l'instant nous ne prendrons qu'un exemple permettant de comparer les possibilités des rayons X dans l'étude des divers défauts des solides.

5. – Un exemple d'étude de défauts: le fluorure de lithium irradié aux neutrons (Lambert et Guinier [13]).

A la suite des chocs des particules primaires et secondaires sur les atomes, il se produit des *lacunes* et des *interstitiels* et l'on sait que les théories des effets de l'irradiation reposent sur les propriétés de ces défauts (par exemple, pour la résistivité des métaux irradiés). Quand l'irradiation est faible, les diagrammes des rayons X ne décèlent pas d'irrégularités dans le réseau: on détecte seulement une légère augmentation du paramètre cristallin mais les raies de diffraction restent fines. Pourtant déjà à ce stade, le cristal est devenu

noir, il y a donc un grand nombre de points de perturbations modifiant considérablement l'absorption de la lumière, mais les rayons X y sont pratiquement insensibles.

Per contre, pour les très hautes doses d'irradiation ($> 5 \cdot 10^{17}$ neutrons/cm^2), on observe sur les diagrammes des rayons X plusieurs sortes d'anomalies. D'abord la largeur des raies augmente considérablement; ensuite il y a diffusion aux très faibles angles. Enfin dans l'espace réciproque des régions de diffusion apparaissent, localisées le long des rangées ⟨100⟩ (et aussi parfois ⟨111⟩). Sans entrer dans le détail de l'interprétation de ces effets, disons seulement que la diffusion aux petits angles est le signe de la présence de cavités submicroscopiques de l'ordre de 20 Å de rayon et que les traînées le long des axes ⟨100⟩ sont produites par des irrégularités dans les plans réticulaires {100}. Ceux-ci, dans des zônes d'une centaine d'angströms de diamètre, auraient un espacement plus grand que la normale et une composition différente de FLi. On est conduit à penser que les lacunes se sont agglomérées pour donner les cavités et que les interstitiels (ici probablement des atomes de lithium) forment des plaquettes d'une couche atomique intercalaire entre deux plans (100) consécutifs. Les rayons X permettent de préciser la taille de ces défauts et de suivre leur évolution au cours de traitements thermiques de « guérison » du cristal.

Ainsi d'une part les méthodes des rayons X apparaissent bien moins sensibles que d'autres méthodes physiques (électriques ou optiques). Par contre les rayons X ont permis de découvrir le phénomène de la coalescence des défauts qui avait été soupçonné mais non encore démontré.

Un autre point à signaler, car il est d'application générale, c'est que les irrégularités les plus aisément décelables sont celles qui ne détruisent pas complètement la périodicité du réseau, par exemple les défauts planaires ou linéaires. En effet, la diffusion étant concentrée dans d'étroites régions de l'espace réciproque, elle est d'intensité relativement beaucoup plus forte que si des désordres atomiques de même amplitudes n'étaient pas coordonnées dans le réseau.

REFERENCES

[1] A. Guinier: *Théorie et Technique de la Radiocristallographie* (Paris, 1956).
[2] R. W. James: *The Crystalline State* (London, 1948).
[3] A. H. Geisler et J. H. Hill: *Acta Cryst.*, **1**, 238 (1948).
[4] K. Huang: *Proc. Roy. Soc.*, A **190**, 102 (1947).
[5] B. E. Warren, B. L. Averbach et B. W. Roberts: *Journ. Appl. Phys.*, **122**, 1493 (1951).
[6] B. S. Borie: *Acta Cryst.*, **10**, 89 (1957).
[7] H. Kunkazi: *Phys. Chem. Solids*, **2**, 24, 107 (1957).
[8] A. J. C. Wilson: *Nuovo Cimento*, **1**, 277 (1955).

[9] T. Suzuki et B. T. M. Willis: *Nature*, **177**, 712 (1956).
[10] B. T. M. Willis: *Proc. Roy. Soc.*, A **239**, 192 (1957).
[11] J. Blin: *Report of the Conference on Defects in Crystalline Solids*, The Physical Society, London (1955), p. 420.
[12] A. Franks et J. Holden: *Nature*, **176**, 1022 (1955).
[13] M. Lambert et A. Guinier: *Compt. Rend.*, **244**, 2623, 2791 (1957).
[14] R. Smoluchowski et Y. Y. Li: *Phys. Rev.*, **94**, 866 (1954).
[15] W. W. Beeman, R. Neynaber, W. Brammer, M. Webb et G. Sharp: *Acta Cryst.*, **10**, 825 (1957).
[16] R. D. Lowde et H. H. Atkinson: *Acta Cryst.*, **10**, 824 (1957).

II.

La structure des solutions solides métalliques.

L'étude de la structure réelle des solutions solides est un bon exemple d'application des méthodes des rayons X pour la détermination des irrégularités cristallines. Les conditions sont favorables si la solution est assez concentrée, car un grand nombre des atomes du cristal entrent en jeu, ce qui donne aux phénomènes de diffusion une intensité suffisante.

On sait maintenant que la solution solide réelle s'écarte sensiblement de la solution solide « idéale » c'est-à-dire où les différents atomes sont, *au hasard* disposés aux nœuds d'un même réseau *régulier*. Ce sont les rayons X qui ont apporté des éclarcissements, à peu près définitifs sur certain points (ordre des atomes); sur d'autres, ils ont soulevé des questions encore non complètement résolues (distorsions du réseau).

Précisons bien notre point de vue. Nous cherchons à faire la géométrie de la structure, en déterminant la position et la nature des atomes. Nous ne nous occupons pas de la description énergétique des atomes, comme on le fait, par exemple, dans la théorie des bandes. Il n'est pas douteux qu'il existe des connexions entre les deux modèles, l'entourage de chaque atome et la distorsion du réseau devant influer sur les niveaux électroniques, mais cette liaison n'est pas faite encore.

1. – Résultats élémentaires des rayons X sur les solutions solides.

Les données des rayons X sur les solutions solides, dans une première approximation, c'est-à-dire en ne tenant compte que des diffractions cristallines normales, sont bien connues. Rappelons-les brièvement, à cause de leur importance.

1) Une solution solide donne un diagramme de diffraction à raies ou taches fines, tout à fait analogue à celui d'un métal pur. Le cristal a donc un

réseau parfaitement déterminé, dont le paramètre peut se mesurer à 1/10 000 près, malgré les différences considérables de diamètre entre les ions constituants (de l'ordre de 10% par exemple).

On admet que les atomes, quelle que soit leur nature, occupent, *en moyenne*, des positions rigoureusement périodiques, les gros atomes étant comprimés et les petits dilatés.

2) Le paramètre varie avec la composition suivant la *loi de Végard*, c'est-à-dire en fonction linéaire de la composition atomique. Cette loi n'est qu'approchée. Théoriquement on a pu la justifier ainsi que les écarts observés (Fournet [1], Friedel [2]).

3) On connaît deux types de solutions solides: la plus commune est la solution *désordonnée*, où tous les sites atomiques jouent exactement le même rôle, les atomes étant disposés au hasard.

Par contre, pour certaines compositions (AB, AB_3), et au-dessous d'une température critique bien définie, la solution peut s'ordonner, les atomes occupant dans la maille élémentaire des positions fixées. Cette *surstructure* se traduit par des raies supplémentaires: ainsi la maille cubique à faces centrées de la solution $AuCu_3$ désordonnée se change en une maille cubique simple à l'état d'ordre, les 4 sites de la maille se différentiant: l'un est occupé par l'or et les 3 autres par le cuivre (Fig. 3). A la température critique (390 °C) les raies de surstructure disparaissent.

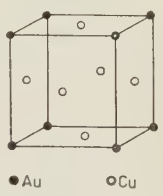


Fig. 3. – Maille de l'alliage $AuCu_3$ à l'état ordonné.

Les raies de surstructure sont seulement visibles si le contraste entre les facteurs de diffusion des atomes A et B est suffisant. Dans le cas d'alliages comme CuZn ou Ni_3Fe, il a fallu des techniques très spéciales pour détecter les raies de surstructure (*Radiocrist.*, p. 394). De tels alliages ne doivent pas être choisis pour l'étude de l'ordre par les rayons X. La diffraction des neutrons est souvent applicable.

2. – Structure réelle des solutions solides.

L'image simple de la solution solide idéale doit être complétée par la description 1) de l'ordonnancement mutuel des atomes qui ne sont ni parfaitement ordonnés ni parfaitement désordonnés, et 2) des déplacements que les atomes subissent à partir des noœuds du réseau moyen idéal.

C'est le premier point qu'il est le plus facile d'étudier avec les rayons X et nous supposerons d'abord, pour l'interprétation des phénomènes de diffraction, qu'il n'y a pas de désordre de déplacement.

3. – Ordre dans les solutions solides.

3·1. *Définitions du degré d'ordre: ordre à grande distance.* – Deux points de vue ont été adoptés; l'un s'applique aux solutions presqu'ordonnées. C'est l'*ordre à grande distance*. Dans le cristal ordonné d'un alliage binaire A-B (p_A et p_B étant les proportions d'atomes A et B), les nœuds sont différenciés en deux types: le groupe 1 recevant les atomes A et le groupe 2 les atomes B. Si l'ordre n'est pas parfait, certains atomes B se trouvent en position 1 et aussi, inversement, certains atomes A en position 2. Soit n la proportion des atomes A en position correcte: n varie de 1 quand l'ordre est parfait, à p_A quand le désordre est complet. Le paramètre d'ordre à grande distance est choisi en fonction de n de façon qu'il varie de 1 à 0, quand on passe de l'ordre au désordre. On pose donc (Lipson [3])

$$S = \frac{n - p_A}{p_B}$$

Du paramètre S dépend l'intensité des raies de surstructure. On montre en effet (*Radiocrist.*, p. 391) que leur facteur de structure est proportionnel à S. Ainsi pour la structure ordonnée $AuCu_3$, ce facteur est

$$(1) \qquad S(f_{Au} - f_{Cu}) .$$

La comparaison sur un même diagramme de l'intensité des raies de surstructure et des raies normales (les raies du réseau désordonné) donne la mesure de S. On trouve que S est nul au-dessus de la température critique T_c, égal à 1 quand on est franchement au-dessous et qu'il prend une valeur un peu inférieure à 1 dans un intervalle d'une vingtaine de degrés au-dessous de T_c: S tend vers 0.8 à T_c pour $AuCu_3$ (Keating et Warren [4]). L'ordre, quand il existe, est donc toujours près d'être réalisé parfaitement. Remarquons que le bloc cristallin dans lequel on définit S doit être tel qu'il y ait diffraction cohérente entre tous les atomes: c'est la base du calcul qui conduit à la relation (1). Autrement dit, des altérations de l'ordre à des distances supérieures au diamètre des blocs mosaïques n'interviennent pas.

3·2. *Ordre à petite distance.* – Soit une solution solide désordonnée à grande distance ($S = 0$). Le grain, défini par la mosaïque, peut-être formé de sous-grains de très petite taille dont chacun est ordonné, mais qui sont décalés les uns par rapport aux autres (« out-of-phase domain »). D'une façon moins précise, on peut dire que les voisins d'un atome de nature donnée ont plus ou moins tendance à s'entourer des mêmes voisins que dans le cristal ordonné.

Il y a donc un certain ordre à petite distance (quelques distances atomiques), mais désordre complet au-delà.

On définit un degré d'ordre pour chacune des couches de voisins, c'est-à-dire les atomes jouant le même rôle cristallographique par rapport à l'atome origine. Soit un alliage, où les A et B sont en proportions p_A et p_B. Considérons les paires d'atomes aux extrémités d'un vecteur du réseau $\boldsymbol{x}_m$, relatif à un atome de la couche m. Nous appelons n_{BA} la proportion de ces couples dont l'atome d'origine est B, et l'atome d'extrémité est A. On pose

$$\alpha_m = 1 - \frac{n_{BA}}{p_A} .$$

L'on vérifie que α_m est égal à 0 dans l'état de désordre complet, car $n_{BA} = p_A$. Pour l'état d'ordre complet, α_m a une valeur déterminée: par exemple 1, si dans cet état, il n'y a que des paires à extrémités semblables pour le vecteur $\boldsymbol{x}_m$. α_m est le *paramètre d'ordre* de la couche m d'atomes.

L'intérêt de ce paramètre est d'une part qu'il mesure le degré d'ordre à petite distance et d'autre part qu'il se relie très simplement au paramètre Φ_m que nous avions introduit dans la théorie générale de la diffraction par un cristal imparfait. On montre (*Radiocrist.*, p. 574) que:

$$\Phi_m = p_A p_B (f_A - f_B)^2 \alpha_m .$$

Nous pouvons donc calculer à partir de α_m l'intensité diffusée: d'après la formule (I.4)

$$I_d(\boldsymbol{s}) = p_A p_B (f_A - f_B)^2 \sum \alpha_m \exp [2\pi i \, \boldsymbol{s} \cdot \boldsymbol{x}_m] , \tag{2}$$

ou comme $\alpha_m = \alpha_{-m}$ et $\alpha_0 = 1$

$$I_d(\boldsymbol{s}) = p_A p_B (f_A - f_B)^2 [1 + 2 \sum \alpha_m \cos 2\pi \, \boldsymbol{s} \cdot \boldsymbol{x}_m] . \tag{2a}$$

Ainsi l'intensité apparaît comme la somme d'un terme constant qui n'est autre que la diffusion de Laue de la solution complètement désordonnée ($\alpha_m = 0$ pour m quelconque, et $\alpha_0 = 1$). L'ordre à petite distance produit des modulations de cette diffusion constante, ces modulations ayant la périodicité du réseau réciproque, car si $\boldsymbol{r}_n$ est un vecteur quelconque du réseau réciproque $\boldsymbol{r}_n \cdot \boldsymbol{x}_m =$ entier pour toutes valeurs de n et m. Donc

$$I_d(\boldsymbol{s} + \boldsymbol{r}_n) = I_d(\boldsymbol{s}) ,$$

car

$$\cos 2\pi \, \boldsymbol{s} \cdot \boldsymbol{x}_m = \cos 2\pi (\boldsymbol{s} + \boldsymbol{r}_n) \cdot \boldsymbol{x}_m .$$

Il suffit donc de déterminer la répartition de la diffusion dans une maille du réseau réciproque; il est avantageux expérimentalement de choisir une maille près de l'origine. En effet 1) les facteurs de diffusion f_A et f_B sont plus élevés, 2) les effets de diffusion par effet de Compton et par agitation thermique sont minimaux autour de l'origine de l'espace réciproque, 3) les effets perturbateurs dûs au déplacement des atomes sont aussi les plus petits possibles dans la première maille, ce qui légitime l'approximation faite en les négligeant.

Si la connaissance de tous les paramètres d'ordre σ_m détermine le diagramme de diffusion, inversement, du diagramme de diffusion on peut déduire tous les paramètres d'ordre (Cowley [5]). Si, en effet, on a établi la carte complète du pouvoir diffusant dans une maille entière du réseau réciproque, par inversion de la somme (2) on obtient:

$$\alpha_m = \int_{\text{maille}} \frac{I_d(s)}{p_A p_B (f_A - f_B)^2} \exp\left[-2\pi i \boldsymbol{s}\cdot\boldsymbol{x}_m\right] \mathrm{d}w_s .$$

Remarquons que les symétries du réseau permettent d'abréger le travail expérimental en n'explorant que la fration de la maille qu'il est nécessaire de connaître pour déduire la maille entière. Des mesures en valeur relative de l'intensité diffusée sont suffisantes puisque l'on sait que α_0 est égal à 1.

Ainsi les rayons X permettent une détermination précise de l'état d'ordre à petite distance. De nombreux exemples ont été traités par Warren et ses élèves [6].

4. – Résultats sur l'ordre à petite distance.

L'essentiel des phénomènes se manifeste sur un exemple très schématique: le réseau linéaire équiatomique AB de période a. La simplicité de ce cas vient de ce que l'on peut caractériser l'ordre par un seul paramètre, la probabilité p qu'une paire de voisins soit dissemblable; $p = \frac{1}{2}$ pour le désordre complet et $p = 1$ si l'ordre est complet (arrangement ABABAB...). p détermine tous les paramètres d'ordre. Pour les premiers voisins $\alpha_1 = 1 - 2p$; pour n-ièmes voisins, α_n est égal à $\alpha_1^n = (1 - 2p)^n$ (*Radiocrist.*, p. 578). La série (2) donnant $I_d(s)$ (*) peut dans ce cas être sommée à l'aide de fonctions élémentaires: le résultat est:

$$I(s) = \frac{(f_A - f_B)^2}{4} \frac{1 - \alpha_1^2}{1 - 2\alpha_1 \cos 2\pi s a_1 + \alpha_1^2} . \tag{3}$$

(*) L'espace réciproque est aussi à une dimension, dans ce cas.

L'intensité de diffraction est représentée sur la Fig. 4 pour différentes valeurs de p. L'intensité diffusée est constante pour le désordre parfait. Quand $p > \frac{1}{2}$, il se produit un maximum juste à l'emplacement des nœuds de surstructure du réseau ordonné. *Ce maximum est d'autant plus accusé et d'autant plus étroit que le degré d'ordre s'accroît.*

Les mêmes caractéristiques se retrouvent sur un cristal réel, comme $AuCu_3$, dont la diffusion au-dessus de la température critique a été étudiée avec une bonne précision par COWLEY [4] (Fig. 5). Cet auteur en a déduit les valeurs du

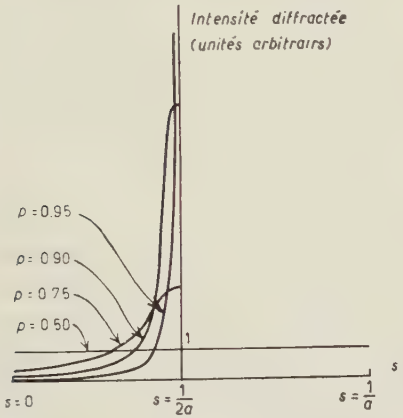


Fig. 4. – Répartition de l'intensité diffractée par un alliage linéaire équiatomique pour différentes valeurs du degré d'ordre, p.

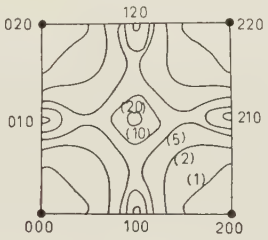


Fig. 5. – Exemple de répartition de la diffusion dans l'espace réciproque correspondant à l'ordre a petite distance: Courbes d'égale intensité dans le plan 100 de $AuCu_3$ (d'après COWLEY).

TABLEAU I. – *Valeurs des paramètres d'ordre à petite distance pour l'alliage* $AuCu_3$ *désordonné* à 405 °C (d'après COWLEY).

Nature des voisins (coordonnées de $\boldsymbol{x}_m$)	α_m pour l'ordre parfait	α_m mesuré à 405 °C
1ers voisins: 110	— 0.333	— 0.152
2èmes » : 200	1.000	0.186
3èmes » : 211	— 0.333	0.009
4èmes » : 220	1.000	0.095
.....	....	....
		$\lvert\alpha_m\rvert < 0.005$

coefficient d'ordre à petite distance de la solution solide en équilibre au-dessus du point critique (Tableau I). Les coefficients trouvés n'ont de valeur nettement différente de zéro que pour les 4 premières couches de voisins et, pour la première couche, α_1 atteint la moitié de la valeur correspondant à l'ordre parfait.

5. – Alliages avec tendance à la ségrégation.

Reprenons l'exemple de l'alliage équiatomique linéaire: mais supposons que la probabilité d'avoir deux voisins dissemblables soit inférieure à $\frac{1}{2}$: cela signifie qu'il y a plus de paires semblables qu'à l'état de désordre parfait; autrement dit, il y a *tendance à ségrégation*. La même formule (3) est encore valable (α_1 est positif): on voit que les maximums se produisent pour les valeurs entières de s, c'est-à-dire à la place des raies de diffraction normales du cristal désordonné. Il y a, en particulier, un maximum d'intensité diffusée autour de l'origine. Ces maximums, comme dans le cas de l'ordre partiel, sont d'autant plus étroits et intenses que la tendance à la ségrégation est plus marquée.

On a observé ces phénomènes dans des solutions solides homogènes au-dessus de la température T_s, où elles atteignent la limite de solubilité. A température inférieure elles se décomposent en deux phases (Fig. 6). Cette tendance à la ségrégation existe déjà dans la phase homogène: il se produit de très petits noyaux où les atomes dissous se rassemblent. C'est la diffusion autour de l'origine qui est le critère le plus net de ces hétérogénéités (Rudman et Averbach [5], Blin, Walker et Guinier [6], Averbach, Flinn et Cohen [17]). La largeur angulaire de cette diffusion permet d'évaluer la taille moyenne de ces noyaux: elle est de l'ordre de 10 Å. Très probablement ces noyaux sont continuellement fluctuants. D'ailleurs il existe déjà de tels noyaux par le simple jeu des fluctuations dues à l'agitation thermique dans une solution parfaitement homogène: mais ils ne donnent naissance qu'à la diffusion continue de Laue. Quand on observe une augmentation de la diffusion au voisinage de l'origine de l'espace réciproque par rapport à la valeur prévue par Laue, c'est que les amas d'atomes dissous d'une dimension donnée sont plus fréquents que ne le voudraient les seules fluctuations statistiques: ils sont donc la preuve d'une interaction entre atomes dissous.

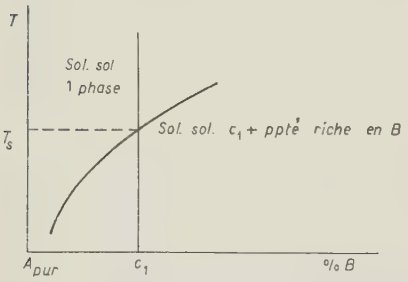


Fig. 6. – Diagramme d'équilibre d'un alliage binaire. Courbe de solubilité.

Ainsi les rayons X ont donné d'intéressantes précisions sur la disposition des atomes au sein d'une solution solide que la théorie élémentaire admet désordonnée. En fait, il n'y a pas d'exemple où la diffusion de Laue soit rigoureusement observée. *L'état le plus fréquent d'une solution solide en équilibre n'est pas le désordre parfait*: il y a, ou tendance à l'ordre ou tendance à la ségrégation.

6. – Solution solide hors d'équilibre.

Les phénomènes d'ordre sont plus prononcés dans les solutions solides en phase unique mais hors d'équilibre, c'est le cas des solutions obtenues par trempe à température ordinaire dans un domaine où l'état d'équilibre est, par exemple, un mélange de deux phases (solution sursaturée) ou encore une solution ordonnée. L'apport des rayons X dans l'étude de la structure de ces états importants a été décisif. Nous consacrerons la prochaine conférence aux troubles d'hétérogénéité avant précipitation. Nous ne nous occuperons ici que des solutions solides pendant la transition désordre → ordre, c'est-à-dire pendant le retour de l'état désordonné métastable à l'état stable ordonné.

Quand on trempe un alliage $AuCu_3$ à partir de l'état désordonné, cet état se maintient à la température ambiante. Si on élève la température (vers 200 ou 300 °C), les observations anciennes de Jones et Sykes [11] avaient montré que les raies de surstructure, caractéristiques de l'ordre, apparaissent progressivement pendant le revenu et qu'elles commencent par être très floues. D'où l'idée d'une structure ordonnée dans de petits domaines, l'ordre n'étant pas cohérent dans les domaines contigus. La Fig. 7 donne une représentation de ces domaines décalés (out-of-step domains) dans le cas d'une rangée linéaire équiatomique. Les domaines sont séparés par des fautes, ici deux atomes semblables. L'origine de ces domaines est facile à justifier. Supposons que l'ordonnancement du réseau débute simultanément en différents points, et que les domaines ordonnés croissent, il n'y a aucune raison pour que les atomes se trouvent partout sur le même réseau, donc il y aura des fautes aux points de jonction des domaines. En supposant que ces domaines diffractent indépendamment, on explique l'élargissement des raies de surstructure par la même effet que pour les raies d'un très petit cristal (*Radiocrist.*, p. 462). Les raies normales, puisque tous les domaines jouent le même rôle, restent fines. Cette interprétation est trop simple pour rendre compte de la répartition de l'intensité diffusée par un cristal unique: la diffusion est toujours concentrée autour de la positions des nœuds de surstructure, mais la tache de diffusion paraît divisée en un certain nombre de satellites du nœud. On pourrait, par la méthode générale, calculer les coefficients d'ordre à petite distance, mais on en tirerait difficilement une image concrète de la structure. Il est donc utile d'essayer des modèles. Ainsi on trouve que les satellites peuvent être expliqués par la présence de domaines ordonnés et décalés, mais à condition de préciser les limites de ces domaines (Wilson [11]), Guinier

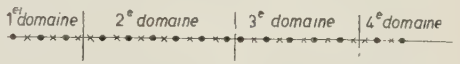


Fig. 7. – Domaine « out-of-phase » d'un alliage linéaire.

et GRIFFOUL [12]) ; ceux-ci seraient limités par des plans de défauts parallèles à {100} et assez régulièrement espacés. On n'a pas encore pu justifier les raisons d'existence de tels domaines.

7. – Désordre de déplacement dans les solutions solides.

Nous avons jusqu'ici admis que les atomes restaient aux nœuds du réseau de la solution solide. Les rayons X donnent alors une idée statistique précise de l'entourage des atomes. On peut à partir de ces résultats faire des calculs thermodynamiques sur les alliages. La théorie la plus simple est la théorie quasi-chimique, où l'on admet que l'énergie de l'ensemble est la somme de l'énergie mutuelle de tous les couples d'atomes. Généralement on considère que seuls les couples d'atomes voisins ont une énergie d'interaction notable. FOURNET [13] a considéré les deux premières couches de voisins et, en introduisant seulement deux constantes, il a pu reproduire les résultats expérimentaux de Cowley sur $AuCu_3$. Mais il est des cas où il y a contradiction entre les données des rayons X et de la thermodynamique. Ainsi l'or et le nickel donnent à haute température une suite continue de solutions solides; au dessous de 800 °C, il y a séparation de la phase unique en deux solutions solides, l'une riche en Au et l'autre en Ni. Or à haute température, les rayons X montrent qu'il y a tendance à l'ordre et non pas tendance à la ségrégation (AVERBACH, FLINN, COHEN [8]). On est conduit à penser que l'équilibre n'est pas seulement déterminé par la nature des paires d'atomes voisins mais aussi par l'énergie élastique provenant des distorsions du réseau. En effet l'écart entre les diamètres d'atomes est ici considérable (15 %) et il est très probable que les atomes sont notablement déplacés en dehors des nœuds du réseau idéal moyen et il serait important pour établir une théorie complète de pouvoir déterminer ces distorsions.

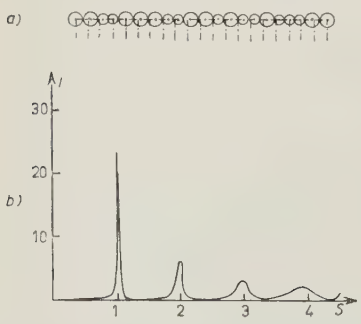


Fig. 8. – *a*) Empilement compact de deux sortes d'atomes de diamètres differents. *b*) Répartition de l'intensité diffractée par l'arrangement *a*).

Quelle réponse peuvent donner les rayons X à ce problème? Reprenons l'exemple schématique de l'alliage linéaire équiatomique pour étudier les effets des déplacements d'atomes.

Une première idée pour arranger les atomes A et B est de les empiler les uns contre les autres en leur conservant le diamètre qu'ils ont dans les métaux purs (Fig. 8). Un tel arrangement est un exemple des imperfections de 2-ème genre décrites (I, Sect. 2·2). En effet, l'ordre à grande distance est altéré,

bien plus que l'ordre à petite distance. La caractéristique de la figure de diffraction donnée par ce genre d'arrangement est que les raies de diffraction ont une largeur croissante avec l'ordre de la réflexion : les nœuds éloignés du centre de l'espace réciproque sont remplacés par des taches de plus en plus élargies et de moins en moins intenses. Or, les alliages comme Au-Cu et Au-Ni, dont les deux atomes présentent une grande différence de volume, donnent toujours des raies de diffraction fines, *même pour les grands indices*. On en déduit que les distorsions ne detruisent pas un réseau moyen, dont les atomes ne s'éloignent que peu.

Notre modèle linéaire ne convient donc pas. Il y a d'ailleurs une différence essentielle entre le cristal et le réseau linéaire. Il n'est pas possible de faire dans l'espace un empilement compact de sphères de diamètres différents. Un réseau régulier ne serait obtenu que si les deux atomes différents prenaient un rayon commun intermédiaire, par déformation élastique. C'est ce modèle qui est probablement plus près de la structure réelle. Mais comme les atomes gardent des rayons un peu différents, le réseau moyen est distordu.

Warren et coll. [4] dans une première approximation, n'ont considéré que les couples d'atomes voisins et ils ont admis, pour une paire, trois distances suivant que c'était une paire AA, AB ou BB. L'effet de ces distorsions est une diffusion qui peut se calculer à partir d'un seul paramètre, déterminé par la différence des rayons $r_A - r_B$, de la proportion des atomes A et B, de leurs facteurs de diffusion f_A et f_B et du paramètre d'ordre α_1. Comme toute diffusion due au désordre des atomes elle est d'intensité croissante avec la distance au centre du réseau réciproque. La superposition de cette diffusion à celle qui provient de l'ordre partiel donne une répartition qui n'est pas symétrique dans la maille du réseau réciproque et qui ne se reproduit plus identiquement de maille en maille. C'est bien conforme aux observations.

En appliquant la théorie de Warren aux solutions solides AuNi, Averbach et coll. [10] ont déduit les distances Au-Au et Ni-Ni dans la solution solide. Pour faire le calcul, ces auteurs admettent que les variations relatives des rayons des atomes par rapport aux rayons moyens déterminés par le paramètre du réseau varient linéairement avec la concentration de l'atome considéré. Il est difficile de juger quelle est la valeur du résultat. Ce qu'on peut dire, c'est qu'il est possible de trouver des paramètres tels que la formule de Warren rende compte des courbes expérimentales. Mais il est certain que cette théorie repose sur des bases trop simplifiées.

Le calcul a été perfectionné par Borie [14] en tenant compte des actions à plus longue distance : celles-ci sont calculées par la théorie de l'élasticité.

Borie a prévu, outre la diffusion trouvée par Warren, deux autres effets :

1) Une décroissance de l'intensité de réflexion analogue à celle due à l'agitation thermique : la mesure expérimentale des intensités donne la valeur de la moyenne quadratique des déplacements d'atomes.

2) Une diffusion autour des nœuds analogue à celle que HUANG avait déjà prévue dans les solutions solides de substitution, comportant deux larges maximums.

Ces différents effets peuvent se calculer à partir d'un seul « paramètre de distorsion » et BORIE a vérifié sur Cu_3Au que les mesures conduisaient bien à partir des différents effets expérimentaux à la même valeur de ce paramètre. Mais l'accord avec l'expérience n'est pas encore parfait.

Si la distorsion du réseau est prouvée expérimentalement, il y a encore des améliorations à attendre des théories qui permettront de préciser les renseignements encore peu précis que nous possédons sur la structure de la solution solide réelle.

REFERENCES

[1] G. FOURNET: *Journ. Phys. Rad.*, **14**, 374 (1953).
[2] J. FRIEDEL: *Phil. Mag.*, **46**, 514 (1955).
[3] D. T. KEATING et B. E. WARREN: *Journ. Appl. Phys.*, **22**, 286 (1951).
[4] J. M. COWLEY: *Journ. Appl. Phys.*, **21**, 24 (1950).
[5] B. E. WARREN, B. L. AVERBACH et B. W. ROBERTS: *Journ. Appl. Phys.*, **22**, 1493 (1951); B. E. WARREN et B. L. AVERBACH: *Modern Research Technics in Physical Metallurgy, A.S.M.*, 93 (1953).
[6] P. S. RUDMAN et B. L. AVERBACH: *Acta Met.*, **2**, 576 (1954).
[7] J. BLIN, C. B. WALKER et A. GUINIER: *Compt. Rend.*, **235**, 254 (1952).
[8] B. L. AVERBACH, P. A. FLINN et M. COHEN: *Acta Met.*, **2**, 92 (1954).
[9] F. W. JONES et G. SYKES: *Proc. Roy. Soc. Lond.*, A **166**, 97 (1938).
[10] G. FOURNET: *Journ. Phys. Rad.*, **13**, 14 (1952).
[11] A. GUINIER et R. GRIFFOUL: *Acta Cryst.*, **1**, 188 (1948).
[12] B. S. BORIE: *Acta Cryst.*, **10**, 89 (1957).
[13] H. LIPSON: *Progress in Metal Physics*, Vol. **2** (New York, 1950), p. 1.

III.

La pré-précipitation dans les solutions solides sursaturées.

Un des exemples les plus typiques d'application des méthodes des rayons X pour l'étude des irrégularités cristallines est fourni par les solutions solides sursaturées pendant le processus de précipitation.

On sait depuis les premiers travaux sur AlCu (GUINIER [1], PRESTON [2]) que les diagrammes des alliages durcissants mettent en évidence, au premier stade du durcissement, des phénomènes de diffusion anormale, traînées ou taches diffuses qui se superposent au diagramme normal de la solution solide homogène. Depuis, ces diffusions anormales ont été étudiées, tant sur Al-Cu que sur d'autres alliages. Nous allons montrer comment on les interprète.

Certes bien des détails ne sont pas encore précisés, mais ce qui est le plus important, c'est que les rayons X ont révélé l'essentiel d'un genre d'imperfection du réseau qui joue un grand rôle dans l'explication du durcissement des alliages.

1. – Insuffisance de la théorie de la précipitation.

Une des premières théories du durcissement structural a été celle de MÉRICA (1920). Le cristal de solution solide sursaturée devient un cristal de solution solide à la concentration de l'équilibre sans changer ni de forme, ni de structure; les atomes expulsés se rassemblent sous forme de très petits grains du précipité qui doit apparaître d'après les diagramme d'équilibre. Mais les grains inclus dans la matrice sont si petits qu'ils ne sont pas perceptibles à l'examen microscopique. Ils augmentent les qualités mécaniques de l'alliage en bloquant les glissements qui se produiraient dans la solution solide homogène. Si l'on adopte cette théorie, on prévoit que les grains submicroscopiques de précipité donnent un diagramme de diffraction X mais dont les taches sont élargies à cause de l'effet de taille ordinaire pour les cristaux. Ce seraient ces taches qui constitueraient les diffusions anormales observées.

Mais cette explication simple n'est pas acceptable, parce que, quelle que soit la forme et la taille d'un cristal, la tache de diffraction est toujours centrée autour du nœud du réseau réciproque du cristal. Or beaucoup des taches observées sont indiscutablement en dehors des taches du précipité d'équilibre. A. GEISSLER [3] a alors essayé de perfectionner cette théorie: 1) en supposant que le précipité formé en grains submicroscopiques pouvait avoir une structure différente de celle du ou des précipités connus. 2) En supposant que les très petits précipités pouvaient avoir un réseau fortement distordu.

Ces hypothèses, d'ailleurs très justifiées, donnent à l'interprétation une grande latitude. Néanmoins certains faits observés ne peuvent entrer dans ce cadre. Citons le plus important, dans le cas de Al-Cu. La diffusion caractéristique du durcissement est constituée par des traînées le long des rangées $\langle 100 \rangle$. Elles pourraient provenir de grains cristallins en forme de disques minces parallèles aux plans $\{100\}$. Mais la largeur des traînées dans l'espace réciproque, en particulier autour du centre, est telle que l'épaisseur du cristal devrait être de l'ordre de 1 à 2 plans atomiques. Est-il possible de parler d'une structure cristalline définie en de si faibles épaisseurs? D'autre part, quel que soit le cristal, une tache de diffraction doit toujours être centrosymétrique. Or les traînées n'ont pas de centre de symétrie. Enfin dans le cas de Al Cu et dans beaucoup d'autres cas, les régions de diffusion semblent beaucoup plus étroitement rattachées aux nœuds de la solution solide qu'aux nœuds d'un autre cristal.

Il semble donc que la théorie de la précipitation fine n'est pas un bon point de départ pour l'interprétation des diagrammes. Il y a, d'autre part, une autre raison, indépendante des rayons X, de chercher une autre explication. C'est la distinction entre les durcissements « à chaud » et « à froid ». Quand on effectue la décomposition de la solution solide AlCu, par exemple à 200 °C, il est indubitable qu'il y a vraie précipitation. Le précipité, ici la phase Al_2Cu θ', se voit au microscope électronique et donne un bon diagramme de rayons X: quelquefois les taches de diffractions sont un peu élargies par effet de taille.

Or quand on effectue le durcissement à basse température (< 100 °C), beaucoup de propriétés physiques ont une évolution différente; c'est le cas, par exemple, de la résistance électrique, de la dilatation, de la cinétique du durcissement... D'où l'idée que le durcissement à froid est lié à une *structure atomique différente*, distincte de la précipitation normale, idée qui nous amène à la théorie des zones (Guinier [4]).

2. – Les zones.

Nous continuons à admettre qu'il y a ségrégation des atomes dissous. La « zone », où il y a ségrégation, est très petite puisqu'elle est invisible au microscope, même électronique (*). Ce que nous introduisons de nouveau, c'est l'idée que la zone n'a pas de structure cristalline bien définie: c'est ce qui la distingue d'un précipité ordinaire.

2·1. *Les zones sans distorsion.* – La plus simple hypothèse serait que les atomes étrangers se rassemblent en noyaux sans quitter les nœuds de la solution solide, ou matrice. Cherchons quelle serait la figure de diffraction de ces zones, supposées toutes identiques et dispersées dans le cristal au hasard. On démontre sans peine (*Radiocrist.*, p. 465) que cette figure est constituée par la répétition autour de chaque nœud du réseau réciproque de la matrice d'une tache de diffusion qui est la figure de diffraction du volume occupé par la zone; son intensité est proportionnelle au carré de la différence des facteurs de diffusion des atomes dissous et solvant.

On a observé des figures de diffraction de ce type, en particulier dans le cas des alliages Al-Ag et Al-Zn (Fig. 9). La diffusion est concentrée dans l'espace réciproque sur une couche sphérique dont le diamètre et l'épaisseur varient avec les conditions de vieillissement; ces petites sphères sont centrées

(*) Toutefois par transmission à travers une feuille d'alliage amincie, Castaing [5] a pu mettre en évidence des zônes dans AlCu d'un diamètre inférieur à 100 Å.

sur tous les nœuds de la matrice et en particulier autour de l'origine de l'espace réciproque. La structure de la zone correspondante peut être précisée, si on admet que celle-ci est centrosymétrique (voir I, Sect. 3·4 et (WALKER et GUINIER [6]).

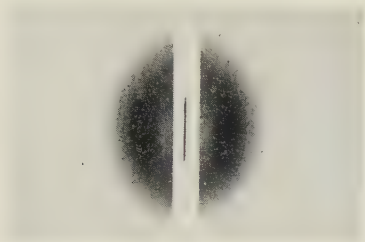


Fig. 9. – Anneau de diffusion entourant le centre du diagramme d'un alliage Al-Ag après vieillissement. Le rayon moyen de l'anneau correspond à un angle de diffusion de 1° environ (radiation Cu K_α).

On peut trouver la répartition de la densité électronique dans la zone par inversion de Fourier de l'amplitude diffusée autour du centre. On peut dans ce cas égaler l'amplitude à $\sqrt{I}$, I étant l'intensité diffusée (*Radiocrist.*, p. 585). Le résultat est la courbe de la Fig. 10. Elle présente une partie positive et une partie négative. Mais il est permis de lui ajouter une densité constante égale à la densité électronique de la solution solide homogène. On peut alors l'interpréter en supposant que le rassemblement des atomes d'argent se fait par diffusion « uphill » à courte distance; la diffusion normale « downhill » étant très lente. Le noyau enrichi en Ag est entouré de l'auréole d'où proviennent les atomes d'Ag ayant migré vers le centre de cette auréole, qui reste donc appauvrie: le tout est entouré de la solution non décomposée. La grandeur de la zone est déterminée directement par l'expérience: par contre, comme on ne connaît pas le nombre des zones, il n'est pas possible de « mettre à l'échelle » la concentration d'atomes qui reste indéterminée en valeur absolue. Toutefois GEROLD [7] a montré qu'un paramètre, qu'il appelle le « degré d'hétérogénéité » peut être calculé sans aucune hypothèse *à priori*.

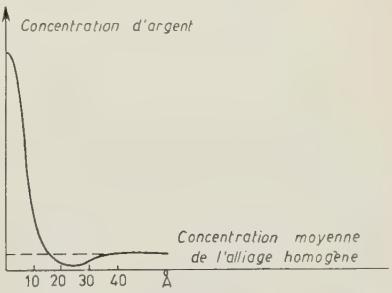


Fig. 10. – Concentration d'argent dans la zone sphérique, déduite par transformation de Fourier de l'intensité mesurée dans la Fig. 9.

Mais ce schéma de zones sans distorsions n'est pas capable de rendre compte des phénomènes observés dans le cas de Al-Cu car il n'est pas possible d'expliquer l'allure très dissymétrique des traînées de diffusion autour des nœuds de la matrice. De même on ne pourrait expliquer le diagramme de l'alliage ternaire Al-Mg-Si qui comporte des diffusions anormales alors que le contraste entre les facteurs de diffusion des atomes est si faible que les rayons X ne sont pas sensibles aux substitutions de l'un des atomes Al, par Mg ou Si. Si le schéma de la zone sans distorsion avait été suffisant dans le cas de Al-Ag et Al-Zn, c'est que les différences de taille entre atomes étaient relativement petites.

2.2. *Les zones avec distorsions.* – Dans le cas général, un rassemblement d'atomes dissous produit inévitablement de fortes distorsions, c'est-à-dire que tous les atomes de la zone et aux alentours sont déplacés hors des nœuds du réseau. La caractéristique de la structure de la zone est qu'elle n'atteint pas un arrangement régulier: nous ferons l'hypothèse qu'elle garde la structure de la matrice plus ou moins distordue. Les distorsions peuvent varier de la périphérie au centre.

En fait la zone n'a pas de limites bien définies et c'est là un autre de ses caractères, il y a parfaite cohérence dans toutes les dimensions avec la matrice environnante. Une telle structure, très influencée par le réseau de la matrice, ne peut exister que si la zone est très petite: l'ordre de grandeur des zones observées est de l'ordre de 100 Å ou moins. Leur forme extérieure varie suivant le système: dans le cas de Al-Cu, où les zones ont été d'abord trouvées, ce sont des plaquettes minces parallèles aux plans (100). Elles sont appelées « Zones Guinier-Preston » ou « Zones G.P. ». Nous emploierons le mot *zone* dans un sens plus général pour des objets d'autres formes.

Nous allons voir maintenant comment à l'aide de la notion de zone on peut interpréter les diffusion observées dans le cas de l'alliage Al-Cu au cours de son vieillissement à froid et quels résultats quantitatifs donnent les rayons X sur la structure des zones.

3. – Interprétation des diffusions données par l'alliage Al-Cu pendant le durcissement.

Nous prendrons comme exemple l'alliage Al-Cu à 4% de Cu. Il est en solution solide homogène au-dessus de 425 °C. Par trempe, nous obtenons à l'ambiante une solution solide métastable sursaturée que nous faisons évoluer vers l'état d'équilibre par des revenus dont on fait varier la durée et la temperature. L'évolution est très complexe suivant la nature de ce revenu.

Nous considérerons uniquement un cas typique du premier stade de structure, par exemple un alliage revenu à 100 °C pendant une dizaine d'heures (Graf [11*b*]). Le diagramme d'un cristal unique oscillant autour d'un axe [100] obtenu avec un rayonnement strictement monochromatique est représenté en Fig. 11.

La caractéristique essentielle est que la diffusion est limitée aux rangées ⟨100⟩. D'où la conclusion immédiate que le cristal contient des défauts de périodicité parallèles aux plans {100}. Il y a, d'autre part, nécessairement, désordre de substitution puisque l'intensité de la diffusion ne s'annule pas au centre de l'espace réciproque. D'où l'idée que les zones sont dues au rassemblement d'atomes de cuivre par taches sur certains plans (100). Par raison de symétrie, il y a trois systèmes de zones sur les trois plans de la forme {100}.

Fig. 11. – Diagramme d'un cristal unique Al-Cu après vieillissement à 100 °C. Axe [100] vertical. Oscillation de 15°; rayonnement Mo K_α.

On aboutit à la transposition à très petite échelle de la disposition dans la matrice des précipités de la phase Al_2Cu que nous révèle le microscope (structure dite de Widmanstätten) (Fig. 12).

On peut facilement trouver le diamètre des zones. En effet si les zones étaient infiniment étendues la diffusion serait rigoureusement concentrée sur la rangée. De même que, pour le cristal tridimensionnel le nœud s'élargit en raison inverse

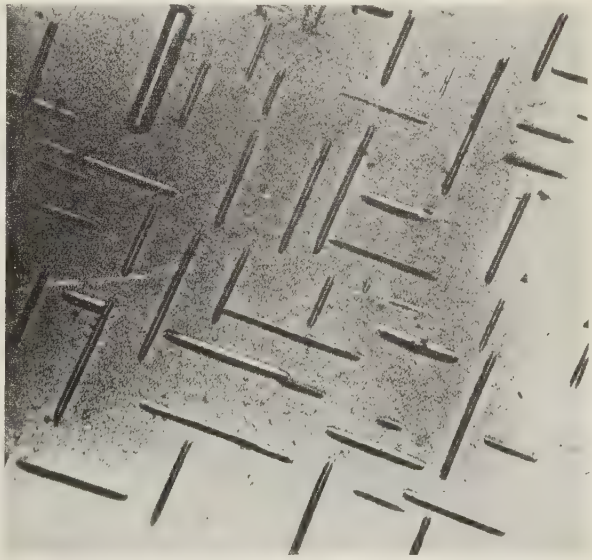


Fig. 12. – Précipités en structure de Widmanstätten. Alliage Al-Cu vieilli a 250 °C; micrographie électronique.

de la taille du cristal, de même la section de la région de diffusion le long de la rangée croît quand le diamètre de la zone diminue (*Radiocrist.*, p. 532). Approximativement, la largeur à mi-intensité comptée en prenant comme unité la maille du réseau réciproque est égale à $1/N$, si N est le nombre de mailles dans le diamètre de la zone.

La mesure n'est possible sur le diagramme (Fig. 11) que si le faisceau primaire est ponctuel. Sinon, la traînée s'élargit d'une quantité — bien plus grande — égale au diamètre du faisceau. Comme le faisceau, issu d'un monochromateur a une section en forme d'une fente très fine, en orientant le cristal de façon qu'il y ait une traînée parallèle à la hauteur du faisceau, sa largeur peut être mesurée de façon significative. On trouve ainsi des dimensions variables avec la température de revenu. Pour 100 °C elle est de 100 Å.

Un autre fait est manifeste: les traînées passent rigoureusement par tous les nœuds de la matrice. On en conclut que dans les plans (100) qui sont irrégulièrement arrangés, les atomes forment toujours rigoureusement le même réseau plan. Mais la figure de diffraction d'un réseau plan isolé d'atomes de cuivre (100) est constitué dans l'espace réciproque par l'ensemble des rangées ⟨100⟩, la répartition de l'intensité étant constante le long de ces rangées (ou plus exactement proportionnelle au carré du facteur de diffusion de l'atome). Ce n'est pas ce qui est observé. Mais ces premières constatations qualitatives, vont nous conduire progressivement au modèle qui a été essayé avec succès.

Le fait important sur lequel nous concentrons maintenant l'attention, c'est que les traînées issues d'un nœud ne son dirigées que d'un côté de ce nœud, *vers les grands angles*. Dès 1938, Preston a étudié le cas du réseau linéaire composé d'atomes de facteur de diffusion variant de façon sinusoïdale tels que, d'autre part, les distances entre atomes voisins varient avec la même période, de façon que les atomes les plus diffusants soient les plus rapprochés: on prévoit que chaque nœud normal est accompagné de satellites, et l'intensité de ces satellites est dissymétrique, le plus intense étant celui du côté des grands angles: c'est le contraire si les atomes les plus diffusants sont les plus écartés.

C'est là un résultat général quand il y a corrélation entre désordres de substitution et de déplacement. On peut le justifier de la façon suivante. Une formule très générale donnant l'intensité diffractée par un ensemble de N atomes de facteur de diffusion f est:

$$I(\boldsymbol{s}) = \sum f_n^2 + \sum_{n \neq n'}^{n} \sum^{n'} f_n f_{n'} \cos 2\pi \boldsymbol{s} \cdot \boldsymbol{x}_{nn'} ,$$

$\boldsymbol{x}_{nn'}$ étant le vecteur séparant les deux atomes n et n'. Considérons les couples de proches voisins, qui sont distants normalement de a. Les termes en $\cos 2\pi s x_{nn'}$ donnent un maximum pour $s = 1/a$ (cas du réseau linéaire).

Les paires de voisins un peu plus rapprochées ou un peu plus éloignées

produisent des maximums pour s plus grand ou plus petit que $1/a$. Mais, si, pour les unes, les facteurs de diffusion sont plus élevés le maximum correspondant sera plus intense. On retrouve ainsi les régles trouvées par PRESTON.

Or les atomes de cuivre sont à la fois plus diffusants et plus petits que ceux d'aluminium. On peut imaginer un rapprochement des plans (100) autour du plan enrichi en cuivre. Les phénomènes de diffusion seront plus intenses du côté des grands angles. La dissymétrie de la traînée est donc du côté attendu: elle signifie que les distorsions jouent un rôle prédominant dans la zone et que ces distorsions sont dues à la différence de taille entre les atomes. Ainsi dans le cas de Al-Ag ou Al-Zn, elles sont négligeables puisque les atomes ont sensiblement le même diamètre.

D'après ces considérations qualitatives, nous pouvons maintenant construire un modèle de zones. Puis en confrontant la diffusion calculée et mesurée, il est possible d'arriver à une description quantitative acceptable des zones de Al-Cu. Deux méthodes ont été employées. GEROLD [8] a essayé divers modèles formés d'un plan médian de cuivre et de plans d'aluminium un peu rapprochés de façon à rattraper le réseau régulier à partir de la n-ième couche (Fig. 13). Il a montré qu'un bon accord avec l'expérience est obtenu pour

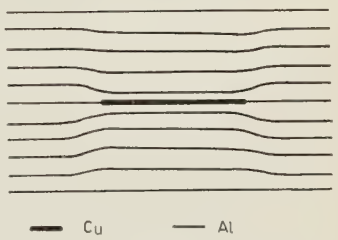


Fig. 13. – Modèle de zone G.P. dans Al-Cu d'après GEROLD.

TABLEAU I.

N°. de la couche	Composition de la couche	Distance du $n^{\text{ième}}$ plan au précédent
0	Cu	—
1	Al	0.9
2	Al	1.0125
3	Al	1.0117
4	Al	1.0108
5	Al	1.0100
6	Al	1.0092
7	Al	1.0083
8	Al	1.0075
9	Al	1.0067
10	Al	1.0058
11	Al	1.0050
12	Al	1.0042
13	Al	1.0033
14	Al	1.0025
15	Al	1,0017
16	Al	1.0008

une zone dont l'influence perturbatrice atteint une quinzaine de couches de chaque côté: les chiffres qu'il a adoptés sont donnés dans le Tableau I.

K. Toman [9] a employé une méthode plus puissante. Il suppose d'abord que la zone a un plan de symétrie dont le plan central est très riche en cuivre. Il suppose ensuite les plans successifs de rang n contenant une proportion de m_n % de cuivre et déplacés d'une quantité *petite* $a\varepsilon_n$, a étant la distance normale des plans (200). Son calcul permet de tirer *directement* de l'expérience les valeurs de m_n et ε_n en fonction de n. Pour les calculs, il faut admettre que m_n décroît rapidement quand n augmente. Le Tableau II donne les valeurs les plus probables d'après Toman.

Tableau II.

N.° du plan n	Proportion du cuivre m_n %	Déplacement relatif ε_n %
0	98	0
1	50	— 3.94
2	39	— 1.54
3	30	— 0.34
4	23	— 0.03
5	17.5	0
6	13	0
7	9	0
8	6	0
9	4	0
10	2.5	0

La principale différence séparant les résultats de Gerold et de Toman est que, pour Toman, il y a jusqu'au sixième plan plus de 10% de cuivre, mais les plans à partir du troisième ne sont pratiquement pas déplacés. Pour Gerold, au contraire, il n'y a qu'un plan de cuivre mais les plans d'aluminium pur sont déplacés substantiellement dans une large bande (15 couches).

Pourtant si l'on compare les écarts entre les valeurs expérimentales et théoriques calculés d'après ces deux modèles, la concordance est assez satisfaisante dans les deux cas. Cela veut dire que les déterminations des structures des zones irrégulières ne pourront vraisemblablement jamais être précises. Il y a à cela plusieurs raisons. D'abord des difficultés expérimentales: en particulier la traînée est masquée au voisinage du nœud par la tache de Bragg et le domaine de diffusion (d'origine thermique) qui l'entoure. Il est difficile d'obtenir les valeurs de l'intensité vraie de la traînée sous la tache de Bragg. D'autre part, il ne faut pas oublier que les données expérimentales, mêmes parfaites et complètes, ne sont pas suffisantes pour restituer une image de l'objet diffractant. Le problème est plus difficile que pour les structures cristal-

lines, car nous ne connaissons pas la composition atomique exacte de la zone, et que les effets de zones, de structure et dimension variables, se superposent. Toutefois il est évident que les modèles de Toman et Gerold, et aussi ceux qui avaient été antérieurement proposés, ont des ressemblances frappantes: leurs points communs peuvent être adoptés avec confiance.

Les rayons X ont ainsi fourni à la physique des solides des éléments très importants sur des structures submicroscopiques qu'aucune autre méthode ne pouvait donner. La zone est une portion du réseau de la matrice dont la composition est anormale et qui est perturbée géométriquement. Mais elle n'a pas de limites bien définies, ni structure propre, elle est enfin cohérente avec la matrice: cet ensemble de caractères différencie la zone et le précipité.

Les zones en forme de plaquettes et les zones quasi-sphériques (type Al-Ag) se retrouvent dans certains autres alliages durcissant (respectivement Cu-Be et Al-Zn). D'autres, au contraire, donnent au premier stade du durcissement des phénomènes de diffusion anormale très différents de ceux que nous avons décrits, mais il est naturel d'attribuer aussi à la présence de zones. Si les observations sont déjà assez complètes, on n'est pas encore aussi avancé dans leur interprétation que dans le cas maintenant classique de Al-Cu et Al-Ag. Les phénomènes les plus curieux ont été trouvés dans les alliages ternaires Al-Mg-Si et Al-Mg-Zn.

4. – Zones à désordre linéaire (Al-Mg-Si).

On observe sur le diagramme d'un cristal de Al-Mg-Si au premier stade du durcissement des lignes continues de diffusion (Lambot [10]) (Fig. 14): il est facile de déterminer qu'elles correspondent à l'intersection avec la sphère d'Ewald (*Radiocrist.*, p. 107) des plans successifs {100} du réseau réciproque. Prenons, par exemple, un cristal orienté avec l'axe [001] vertical, le faisceau de rayons X, horizontal, étant parallèle à [100]. On observe une série de fines lignes verticales d'intensité très faible. L'une passe par le centre; la seconde est à la distance a^* et la troisième à la distance $2a^*$, $a^* = 1/a$ est le paramètre de la maille du réseau réciproque, le réseau cristallin ayant pour paramètre a. En réalité, ces lignes ne sont pas des droites mais des hyperboles très ouvertes et des mesures précises montrent que ce sont les projections des cercles C_0, C_1, C_2, intersection de la sphère d'Ewald avec les plans (010) du réseau réciproque (Fig. 14).

On ne voit pas les lignes horizontales qui, par raison de symétrie, devraient apparaître: cela vient de ce que le faisceau issu du monochromateur est un pinceau d'une hauteur de quelques millimètres. Une ligne de diffusion verticale n'est pas élargie de ce fait, tandis qu'une ligne horizontale aurait une

hauteur égale à la hauteur du faisceau: son intensité devient trop faible pour être décelable.

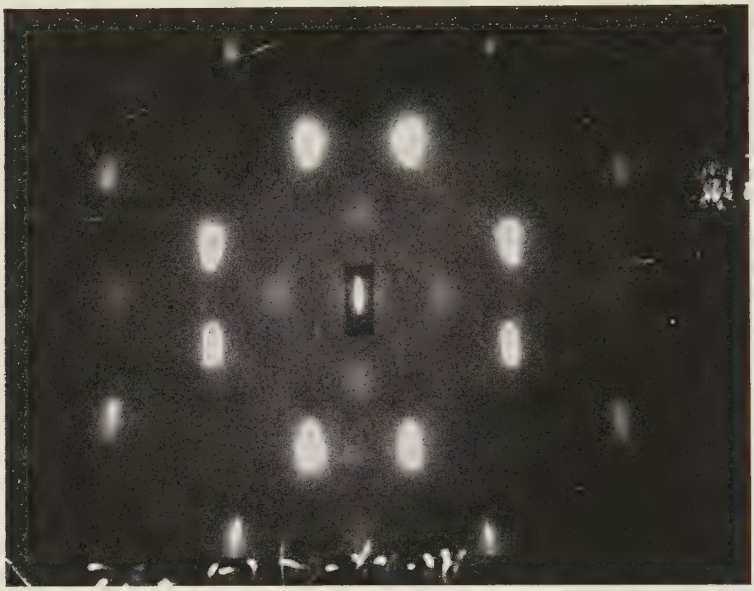


Fig. 14. – Diagramme de diffraction d'un cristal Al-Mg-Si vieilli à 150 °C. Cristal immobile. Axe [001] vertical, cristal irradié suivant [100]. Rayonnement Mo K_α (Lambot).

Les lignes observées sont très étroites, ce qui signifie que la diffusion est rigoureusement concentrée sur les plans {100} du réseau réciproque. Les lignes paraissent, tout au moins aux premiers stades du durcissement, d'intensité continue et lentement décroissante quand on s'éloigne du centre de l'espace réciproque. Si l'on modifie l'orientation du cristal la forme des lignes change comme le prévoit la géométrie mais leur intensité reste à peu près uniforme. On peut donc admettre, *en première approximation*, que les plans {100} de l'espace reciproque sont le siège d'une diffusion à peu près uniforme sur toute leur étendue, sauf toutefois au voisinage du centre. Car l'intensité des lignes est nettement diminuée autour de l'origine. D'autre part, des modulations de l'intensité se produisent quand le revenu de l'alliage est plus poussé: mais nous ne nous occupons que du premier stade. Quelles conclusions pouvons nous tirer sur l'origine de ces diffusions anormales?

Certaines déductions sont immédiates et sûres:

1) Le désordre est linéaire, la périodicité de la zone perturbée est rigoureusement maintenue le long des rangées ⟨100⟩ car la diffusion est contenue dans les plans {100}.

2) Cette périodicité rigoureuse est cohérente sur une grande longueur (> 500 Å), puisque l'épaisseur des plans de diffusion est si faible qu'elle n'est pas mesurable. Enfin l'intensité est uniforme sur les plans, c'est-à-dire que nous observons à peu près la figure de diffraction d'une ligne unique de points, un réseau linéaire régulier de période égale au paramètre de la matrice le long de [100].

Tous les atomes de l'alliage Al-Mg-Si ont des pouvoirs diffusants très voisins, on ne peut donc pas imaginer des zones formées de files d'atomes isolés, puisque le rassemblement du magnésium et du silicium serait invisible aux rayons X. D'ailleurs dans ce cas l'intensité ne diminuerait pas vers le centre de l'espace réciproque. Pourtant la ségrégation de Mg et Si est certainement l'origine des phénomènes. On est conduit à admettre que cette ségrégation provoque un désordre de position dans les zones enrichies en Mg et Si, et que ce désordre laisse intacte la périodicité des atomes le long des axes quaternaires. Ces zones doivent avoir la forme de longues aiguilles, ce qui est confirmé par le fait que les précipités Mg_2Si, quand ils se forment dans l'alliage par des revenus à plus haute température, ont aussi la forme d'aiguilles paral lèles à [100]. Il n'est pas encore possible d'avancer des hypothèses sur la constitution de ces zones: rien ne permet, par exemple, d'évaluer leur diamètre. On peut se demander si la stabilité de ces curieuses zones n'est pas liée à la formation de liaisons d'un caractère non métallique entre Mg et Si.

5. – Zones de Al-Mg-Zn (Graf, [11 *a*]).

Nous citerons encore une autre série d'observations sur des diffusions anormales surtout pour exposer le problème qu'elles posent, car l'interprétation est encore douteuse.

Les observations sont les suivantes:

1) Autour du centre, on trouve un anneau assez semblable à celui que donnent les alliages Al-Zn. Cette diffusion centrale prouve qu'il y a des ségrégations d'atomes de Zn, puisqu'on ne peut rien dire de la répartition du magnésium que les rayons X ne peuvent distinguer de l'aluminium. La structure de ces zones est semblable à celle des zônes de Al-Ag. Les dimensions des zones variant suivant la composition de l'alliage et le traitement de vieillissement: leur ordre de grandeur est de 50 à 100 Å.

2) En plus le diagramme comporte une série de taches diffuses correspondant dans l'espace réciproque à des « boules » d'un rayon de l'ordre de 1/10 de la maille du réseau réciproque (Fig. 15). Certaines de ces boules sont en position simple par rapport aux nœuds de la matrice, elles ressemblent aux

taches apparaissant autour des nœuds d'un alliage à l'état d'ordre partiel. D'où l'idée que dans la zone il y aurait commencement d'ordre entre les atomes Mg, Zn et Al. Ces phénomènes n'existent pas dans le cas de Al-Zn. Cet ordon-

Fig. 15. – Diagramme de diffraction d'un cristal Al-Mg-Zn (GRAF).

nancement serait donc la conséquence d'une interaction entre Mg et Zn. Jusqu'à présent ces idées qualitatives n'ont pas été approfondies, à cause du manque de données complètes et quantitatives sur l'espace réciproque de cet alliage.

Des phénomènes analogues ont été trouvés dans Al-Mg-Cu (BAGARYATSKII [12], LAMBOT [10]).

REFERENCES

[1] A. GUINIER: *Compt. Rend.*, **208**, 1011 (1938).
[2] G. D. PRESTON: *Proc. Roy. Soc. Lond.*, A **167**, 526 (1938).
[3] A. H. GEISSLER et J. K. HILL: *Acta Cryst.*, **1**, 238 (1948).

[4] A. Guinier: *Journ. Met.*, 673 (1956).
[5] R. Castaing et G. Lenoir: *Compt. Rend.*, **239**, 972 (1954).
[6] C. B. Walker et A. Guinier: *Acta Met.*, **1**, 568 (1953).
[7] V. Gerold: *Zeits. f. Metallk.*, **46**, 623 (1955).
[8] V. Gerold: *Zeits. f. Metallk.*, **45**, 599 (1954).
[9] K. Toman: *Acta Cryst.*, **10**, 187 (1957).
[10] H. Lambot: *Rev. Metal.*, **47**, 709 (1950).
[11] R. Graf: *a*) *Compt. Rend.*, **242**, 1311, 2384 (1956); *b*) *Publ. Scient. Techn. Minist. Air*, N°. 315 (1956).
[12] A. Bagaryatskii: *Journ. Techn. Phys.* (*URSS*), **20**, 424 (1950).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Gitterstörungen in dünnen Schichten.

R. Hilsch und W. Martienssen
Universität Göttingen - Göttingen

1. – Einleitung.

Zur Erzeugung von Gitterstörungen in kristallinen Stoffen werden seit Jahren eine Reihe verschiedener Methoden benutzt. Die älteste Methode ist die der plastischen Deformation fester Körper. Daneben ist in den letzten Jahren die Einwirkung energiereicher Strahlen auf den kristallinen Aufbau besonders von technisch wichtigen Materialien sehr ausgiebig untersucht worden. In beiden Fällen hat man es mit Gitterstörungen im eingefrorenen Zustand zu tun: Durch eine Wärmebehandlung läßt sich der Ausgangszustand weitgehend wiederherstellen. Andererseits kann man bei hohen Temperaturen auch Gitterstörungen im thermischen Gleichgewicht erzeugen, man hat eine Gleichgewichtskonzentration an einzelnen Fehlstellen (z.B. Lücken etc.) und außerdem eine Störung des Gitters durch die Temperaturbewegung. In dem vorliegenden Bericht werden einige Ergebnisse zusammengestellt, die mit einer vierten Methode erzielt worden sind: Der Herstellung gestörter kristalliner Schichten durch Aufdampfen bei tiefen Temperaturen.

Mit dieser Methode ist es möglich, außerordentlich hohe Konzentrationen an Gitterstörungen herzustellen, man gelangt zu einem ganz neuartigen Zustand der Materie mit interessanten, veränderten Eigenschaften. Wir bezeichnen diese Methode als «abschreckende Kondensation». Die in Frage stehende Substanz wird im Hochvakuum verdampft und auf einer auf tiefe Temperatur abgekühlten Unterlage kondensiert. Die Atome oder Moleküle werden also aus dem Dampfstrahl plötzlich auf eine sehr tiefe Temperatur abgeschreckt und vermögen sich dann nicht mehr zu einem ungestörten Kristallverband zu ordnen. Die Kondensationswärme wird durch die Wärmeleitung der Unterlage sehr schnell abgeführt. Man gelangt so zu hochgradig fehlgeordneten Schichten mit Dicken bis zu etlichen Zehntel μm. Für die Eigenschaften dieser Schichten spielt ihre Dicke keine wesentliche Rolle.

Ohne Schwierigkeiten lassen sich auch homogen zusammengesetzte Schichten aus mehreren verschiedenen Substanzen herstellen. Haben die einzelnen Komponenten verschiedene Dampfdrucke, so tritt beim langsamen Verdampfen der Mischung eine Fraktionierung auf. Diese läßt sich aber dadurch vermeiden, daß die Mischung nacheinander in vielen kleinen Portionen verdampft wird, von denen jede einzelne nur eine einatomare Schicht liefert.

In dem vorliegenden Bericht behandeln wir folgende Eigenschaften solcher Schichten:

1) Röntgen- und Elektronenbeugung,
2) Optische Absorption,
3) Elektrische Leitfähigkeit und Supraleitung,
4) Fehlordnungsenergie.

2. – Röntgen- und Elektronenbeugung.

Wir beginnen unsere Diskussion mit Röntgenbeugungsaufnahmen an Salzschichten, zunächst der Alkalihalogenidkristalle. Fig. 1, entnommen aus einer Arbeit von Rühl [1], zeigt Debye-Scherrer-Aufnahmen an einer bei 20 °K kondensierten KI-Schicht, Teilbild *a* ist sofort nach der Kondensation bei 20 °K aufgenommen, Teilbild *b* ist im Anschluß an eine Temperung auf 330 °K wiederum bei tiefen Temperaturen gemessen. Die frisch kondensierte Schicht liefert stark verbreiterte Röntgen-Reflexe, als Halbwertsbreite findet man 0.6 : 0.8 Bragg'sche Winkelgrade; nach der Temperung beträgt die Halbwertsbreite nurmehr etwa 0.25 Winkelgrade und entspricht damit schon ungefähr der vorgegebenen Apparatebreite von etwa 0.15°. Ganz entsprechende Ergebnisse findet Rühl bei fast allen anderen Alkalihalogeniden. Man erhält die Interferenzen des normalen bei Zimmertemperatur stabilen Gitters (NaCl-bzw. CsCl-Typ), aber mit stark vergrößerter Halbwertsbreite. Die Schichten sind also als kristallin anzusehen, befinden sich aber in einem stark gestörten Zustand. Man wird es mit vielerlei verschiedenen

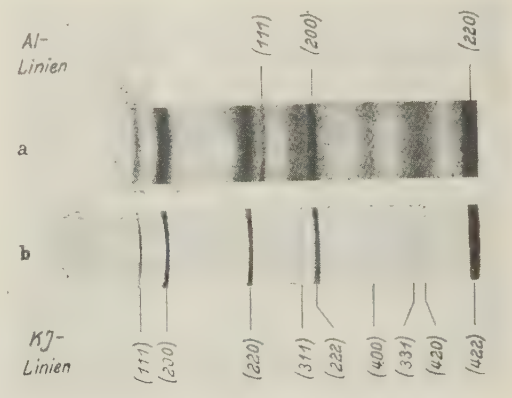


Fig. 1. – Debye-Scherrer-Aufnahmen einer bei 20 °K kondensierten KI-Schicht. *a*) aufgenommen nach der Kondensation bei 20 °K; *b*) nach Temperung bis 330 K aufgenommen bei 90 K. Belichtungszeiten: *a*) 150 min, *b*) 75 min.

Arten von Fehlordnung zu tun haben: Lücken, Lückenaggregate, Versetzungen, Korngrenzen, Verspannung der kristallinen Bereiche und so fort. Deutet man die Verbreiterung der Interferenzen allein mit der geringen Korngröße, so folgen daraus Kristallitgrößen von $100 \div 200$ Å.

R. HILSCH [2] hat ein ebenes Modell angegeben, das ein anschauliches Bild von der Struktur solcher Aufdampfschichten vermittelt. Die Ionen werden durch kurze, zylindrische Magnete dargestellt (Fig. 2). Schwarze Nordpolenden bedeuten I^--Ionen, weisse Südpolenden K^+-Ionen. Das linke Teilbild der Fig. 2 zeigt das Modell einer bei tiefen Temperaturen kondensierten KI-Schicht, entstanden durch Aneinanderwerfen von Modell-Molekülen. Man erkennt deutlich die kleinen, teilweise verspannten kristallinen Körner, voneinander getrennt durch weitgehend fehlgeordnete Bereiche. Das rechte Teilbild der Fig. 2 zeigt die beginnende Kristallisation beim Aufwärmen der Schicht, verwirklicht durch Schüttelbewegung.

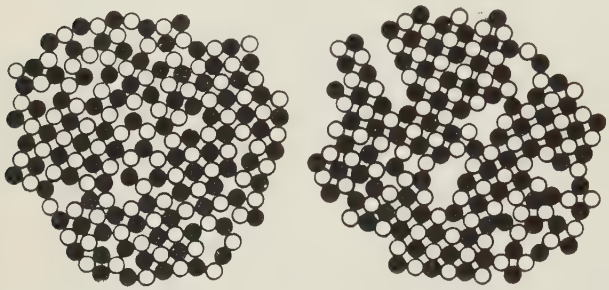


Fig. 2. – Flächenhaftes Modell einer KI-Schicht, dargestellt durch zylindrische Magnete. ● = I^--Ion, ○ = = K^+-Ion. Links: fehlgeordnete Schicht nach der Kondensation bei tiefen Temperaturen; rechts: beginnende Kristallisation beim Aufwärmen (Schüttelbewegung am Modell).

Eine Ausnahme im Verhalten der Alkalihalogenidschichten bildet das LiI. LiI kristallisiert normalerweise in der NaCl-Struktur. Bei der Kondensation bei 20 °K entsteht LiI jedoch in einer neuen, bisher nicht bekannten hexagonalen Modifikation. Beim Erwärmen auf Zimmertemperatur tritt innerhalb von Stunden die Umwandlung in das normale NaCl-Gitter ein. Das Ionenradienverhältnis $r_{\text{Kation}}/r_{\text{Anion}}$ beträgt bei LiI 0.35 und liegt damit in dem Bereich, in dem nach BORN und MAYER [3] die Gitterenergien für die NaCl- und die ZnO-Struktur nahezu gleich sind. Die aus den Abmessungen der Elementarzelle ermittelte röntgenographische Dichte beträgt für die hexagonale Struktur $3.5\ \text{g}\cdot\text{cm}^{-3}$, für die NaCl-Struktur $4.12\ \text{g}\cdot\text{cm}^{-3}$. Die neue Struktur entspricht also einer weniger dichten Packung und ist in der bei tiefer Temperatur kondensierten Schicht offenbar günstiger, da diese Schicht durch den hohen Grad eingefrorener Fehlordnung aufgelockert ist.

Bei der Untersuchung von Schwermetallsalzen findet RÜHL [4] im allgemeinen das gleiche Verhalten wie bei KI: die bei tiefen Temperaturen kondensierten Schichten entstehen in einem gestörten, kristallinen Zustand. Ein

ganz neuer Befund ergibt sich aber bei den Halogeniden des Thalliums, sowie dem Silber- und Kupferjodid. Fig. 3 zeigt Schwärzungskurven von Debye-Scherrer-Aufnahmen an einer aufgedampften TlI-Schicht, Teilbild a wieder sofort anch dem Aufdampfen bei 20 °K, Teilbild *b* nach der Temperung. Die breite Interferenz im Teilbild *a* deutet einen amorphen, flüssigkeitsähnlichen Zustand der frisch aufgedampften Schicht an. Beim Tempern erhält man in dem schmalen Temperaturbereich zwischen 135 und 155 °K den Übergang zu den scharfen Interferenzen des normalen orthorhombischen TlI. Im Gegensatz zu den Ergebnissen bei den Alkalihalogeniden ist zwischen der Lage der Interferenzen im kristallisierten und im frisch aufgedampften Zustand kein Zusammenhang zu erkennen.

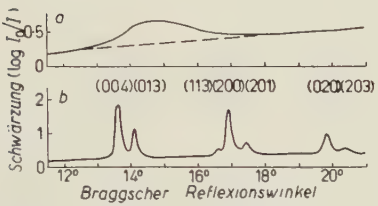


Fig. 3. – Schwärzungskurven von Debye-Scherrer-Diagrammen einer bei 20 °K kondensierten TlI-Schicht. *a*) nach der Kondensation aufgenommen bei 20 °K; *b*) nach Tempern bis 330 °K aufgenommen bei 80 °K.

Eine Ursache für dieses unterschiedliche Verhalten ist in der Polarisierbarkeit der Ionen zu suchen. Die statische Dielektrizitätskonstante der Alkalihalogenide liegt im Bereich zwischen 4.5 und 6.5, die der Tl-Halogenide aber zwischen 30 und 35. Bildet man das Verhältnis zwischen nächsten Ionenabständen im Kristall und dem Ionenabstand im Dampfmolekül, so findet man bei den Alkalisalzen Werte zwischen 1.09 und 1.17, bei den Tl-Salzen dagegen 1.3. Bei der Kondensation der Tl-Halogenide ist also eine viel stärkere Aufweitung der Ionenabstände erforderlich als bei den Alkalihalogeniden. Die weitreichenden heteropolaren Bindungskräfte sind durch die Polarisation stark abgeschirmt. So kondensiert die Schicht bei hinreichend tiefen Temperaturen in einem völlig ungeordneten Zustand (bei rein homöopolar gebundenen Substanzen, z.B. Ge und Si, erhält man sogar noch bei Zimmertemperatur amorphe Aufdampfschichten).

Bei Metallen führt die abschreckende Kondensation im allgemeinen ähnlich wie bei den zuerst genannten Salzen ebenfalls zu gestörten feinkristallinen Schichten. Fig. 4 zeigt als Beispiel Elek-

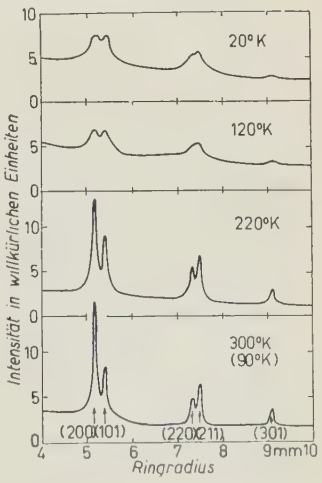


Fig. 4. – Schwärzungskurven von Elektronenbeugungsaufnahmen an einer reinen, bei 20 °K kondensierten Zinn-Schicht. Oberes Teilbild sofort nach der Kondensation, aufgenommen, die folgenden Bilder bei den angegebenen Temperaturen. Aufnahme 4 bei 90 °K nach Tempern auf 300 °K.

tronenbeugungsaufnahmen von W. BUCKEL [5] an einer aufgedampften Zinn-Schicht. Der Lage nach stimmen die Interferenzen der getemperten (unteres Teilbild) und der frisch aufgedampften (oberes Teilbild) Schicht überein. Die Breite der Interferenzlinien ist jedoch nach der abschreckenden Kondensation stark vergrößert und läßt auf Korngrößen von etwa 100 Å schließen.

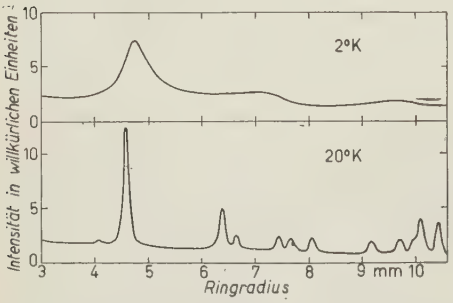


Fig. 5. – Schwärzungskurven von Elektronenbeugungsaufnahmen einer bei 2 °K kondensierten Wismut-Schicht. Oberes Teilbild sofort nach der Kondensation, unteres Teilbild nach Erwärmen auf 20 °K.

Dieses Verhalten ist typisch für eine große Anzahl von Metallen. Zwei Ausnahmen davon stellen Wismut und Gallium dar (BUCKEL [5]). Eine bei 2 °K kondensierte Bi-Schicht entsteht in einem amorphen Zustand, wie die Beugungsaufnahme Fig. 5 zeigt. Beim Erwärmen auf 14 °K ändert sich an dem Beugungsbild nichts. Wenige Grade höher jedoch verschwinden die breiten Beugungsringe und es erscheinen die Interferenzen des normalen Bi. Ähnlich erhalten wir auch beim Ga nach der Kondensation bei 4 °K zunächst breite Interferenzringe, die auf einen flüssigkeitsähnlichen Zustand hindeuten. Beim Tempern auf 20 °K erscheint ein System scharfer Ringe, das jedoch nicht dem normalen Gallium-Gitter zugeordnet werden kann. Erst oberhalb 60 °K erhält man nach einer nochmaligen Umwandlung die Interferenzen des normalen Ga-Gitters. Bei erneutem Abkühlen auf tiefe Temperaturen bleibt dieses Gitter erhalten. Die bisher unbekannte Modifikation des Ga stellt eine Analogie zur hexagonalen Modifikation des LiI dar. Auch hier wird die neue Struktur offenbar begünstigt durch die eingefrorene Fehlordnung in der aufgedampften Schicht. Im thermischen Gleichgewicht ist bei tiefen Temperaturen nur das normale Ga-Gitter stabil. Eine Messung der Wärmetönung bei der Gitterumwandlung (vgl. Abschnitt 5) ergibt, daß es sich bei dem unbekannten Gitter um eine Hochtemperaturmodifikation handelt. Im thermischen Gleichgewicht läßt sich dieses Gitter aber nicht herstellen, da Ga bereits bei 28 °C schmilzt.

Beim Aufdampfen reiner Substanzen gelangt man nur in einigen Sonderfällen zu amorphen Schichten, im allgemeinen entstehen die Schichten in einem zwar gestörten aber noch kristallinen Zustand. Man kann jedoch auch in diesen Fällen die Kristallisation weitgehend behindern, indem man fremde Atome oder Moleküle in geringer Konzentration gleichzeitig in die Schicht hineindampft. Fig. 6 zeigt als Beispiel Photometerkurven von Beugungsaufnahmen an einer Sn-Schicht mit 10 At. % Cu. Die stark verbreiterten Beugungsringe nach der Kondensation weisen auf einen flüssigkeitsähnlichen oder

feinstkristallinen Zustand hin. Bei 30 °K sieht man aus den breiten Interferenzen verhältnismäßig scharfe Linien herauswachsen (vgl. Fig. 4). Hier entstehen offenbar, von wenigen Keimen ausgehend, in der äußerst ungeordneten Umgebung wenige größere Kristallite, im Gegensatz zur Sammelkristallisation aus der schon kristallisierten Umgebung bei reinem Sn. Im unteren Teilbild der Fig. 6 treten zusätzliche Linien einer intermetallischen Verbindung Sn-Cu auf (schräge Pfeile).

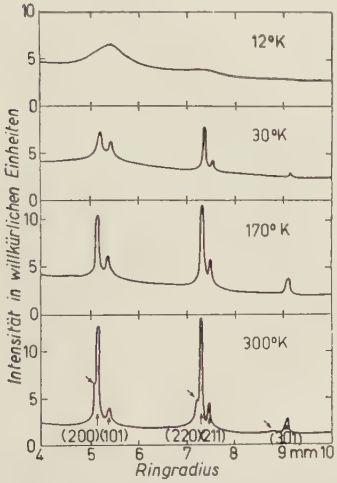


Fig. 6. – Schwärzungskurven von Elektronenbeugungsaufnahmen an einer Aufdampfschicht Sn + 10 At. % Cu, oberes Teilbild aufgenommen nach der Kondensation bei 12 °K, die weiteren Bilder nach Erwärmen auf die angegebene Temperatur.

Entsprechend lassen sich auch Salzschichten, die normalerweise kristallin kondensieren, durch geringe Fremdzusätze in einen amorphen Zustand zwingen. Ein Beispiel ist das System KI-KF, das in Abschnitt 3 behandelt wird. Die Wirksamkeit des Zusatzes läßt sich wieder sehr deutlich im Magnetmodell [2] demonstrieren (Fig. 7): Durch den Zusatz der kleinen Fluor-Ionen erhält man links beim Aufdampfen eine quasiamorphe Schicht, nach dem Tempern (Schütteln) zeigt sich ein ausgeschiedenes KF-Kriställchen in der kristallisierten KI-Umgebung.

Im Beispiel KI + 10 % KF wirkt das Fluorid behindernd auf die Kristallisation des KI ein, da KF mit KI keinen Mischkristall zu bilden vermag. Bei hinreichend tiefen Aufdampftemperaturen erhält man amorphe Schichten. Es ist eine interessante Frage, was beim Zusammendampfen zweier nichtmischkristallbildender Komponenten passiert, wenn die Schicht bei höheren Temperaturen kondensiert wird. QUEISSER [6] has das System NaCl-CsCl mit Hilfe von Röntgenbeugungsauf-

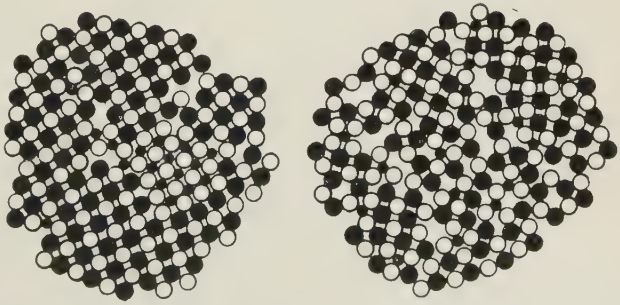


Fig. 7. – Flächenhaftes Modell einer Aufdampfschicht KI + 10 At. % KF. Die Fluorionen werden durch kleine, schwarze Kreisflächen dargestellt (zylindrische Magnete mit kleinerem Durchmesser, aber mit gleichem Moment). Links: stark fehlgeordnete Schicht nach der Kondensation bei tiefen Temperaturen; rechts: Ausscheidung des KF beim Tempern (Schütteln).

nahmen untersucht. Im thermischen Gleichgewicht bilden die beiden Komponenten keine Mischkristalle. Bei Kondensationstemperaturen oberhalb 250 °K erhält man bei starkem Überwiegen einer der beiden Komponenten eine feinkristalline Schicht von NaCl- bzw. CsCl-Kriställchen: Die überwiegende Komponente vermag den Zusatz auszuscheiden, dieser verhindert jedoch die Bildung größerer Kristallite. Bei mittleren Konzentrationen aber können sich die beiden Partner nicht gegenseitig ausscheiden, es bildet sich ein erzwungener Mischkristall. Beim Tempern oberhalb 50 °C zerfällt dieser in die beiden reinen Komponenten. Wird eine Schicht mittlerer Konzentration bei tiefen Temperaturen (20 °K) aufgedampft, so erhält man eine amorphe Struktur; bei dieser tiefen Temperatur reicht die Wärmebewegung zur Bildung der erzwungenen Mischkristalle nicht mehr aus.

3. – Optische Absorption.

In Abschnitt 2 haben wir Beugungsspektren an gestörten Aufdampfschichten behandelt. Sie werden entscheidend beeinflußt durch den Zustand der kristallinen Ordnung über größere Gitterbereiche hin (Fernordnung). Im folgenden soll untersucht werden, in welcher Weise das optische Absorptionsspektrum durch Gitterstörungen verändert wird. Für diesen Zweck erscheinen die Ionen-Kristalle besonders geeignet; ihre Spektren enthalten schmale Absorptionsbanden, die lokalisierten Elektronenübergängen im Kristallgitter zugeordnet werden können. Die Lage und Gestalt dieser Absorptionsbanden in gestörten Schichten liefern Aussagen über die Beschaffenheit der direkten Umgebung der jeweiligen Absorptionsstelle und somit über die Nahordnung in dem gestörten Kristall. Als Gitterstörungen wollen wir dabei zunächst die eingefrorene Fehlordnung in abgeschreckt kondensierten Schichten und dann die durch die thermischen Gitterschwingungen hervorgerufene Unordnung im Kristall betrachten.

Das optische Absorptionsspektrum der Alkalihalogenide beginnt im kurzwelligen Ultraviolett und besteht, wie zuerst HILSCH und POHL [7], sowie SCHNEIDER und O'BRYAN [8] gezeigt haben, aus einer Reihe mehr oder weniger gut voneinander getrennter Banden, von denen die langwelligsten mit einem Elektronenübergang von einem Halogenion zu einem direkt benachbarten Alkaliion gedeutet worden sind. FRENKEL [9] und WANNIER [10] haben als erste diese Banden als Excitonenbanden erklärt. Neuerdings haben PHILIPP und TAFT [11] mit Hilfe lichtelektrischer Messungen auch die langwellige Grenze für optische Band-Band-Übergänge im Spektrum der Alkalihalogenide festlegen können. Die Spektren der Silbersalze haben nach neuen Messungen von TUTIHASI [12] und OKAMOTO [13] eine ganz ähnliche Gestalt wie die der Alkalihalogenide, nur scheint hier im Gegensatz zu diesen die Band-Band-Absorption bereits auf der langwelligen Seite der Excitonenbanden einzusetzen.

F. Fischer [14] hat den Einfluß der abschreckenden Kondensation auf das Absorptionsspektrum aufgedampfter KI-Schichten untersucht (Fig. 8). Im oberen Teilbild erkennt man die Spektren einer Reihe reiner KI-Schichten,

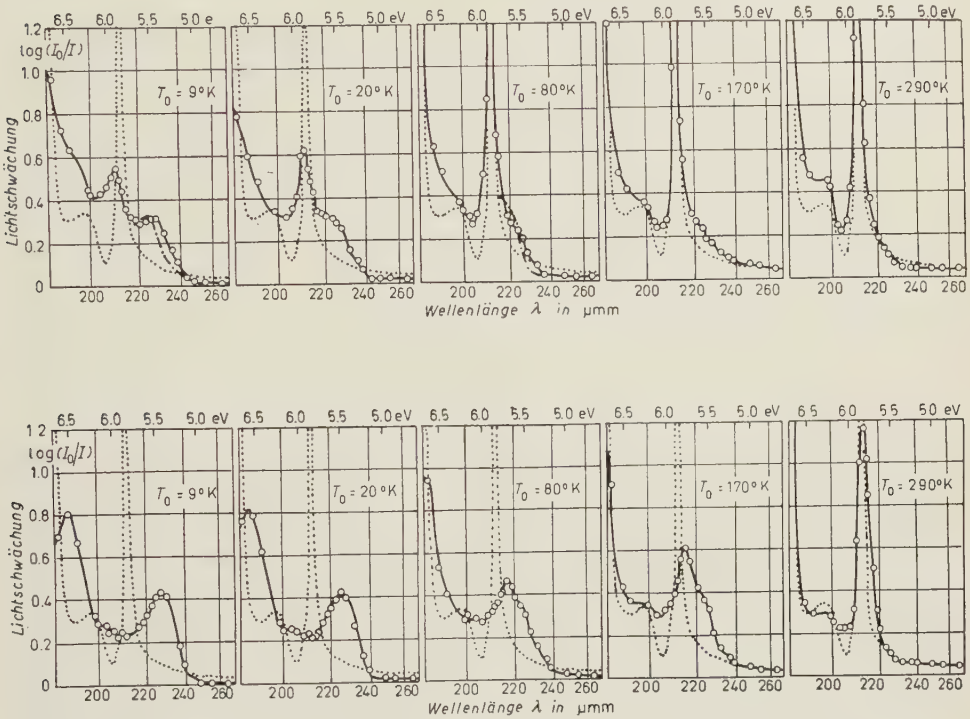


Fig. 8. – Obere Reihe: Absorptionsspektren von KI-Schichten, kondensiert bei der Temperatur T_0, gemessen bei 20 °K bzw. 9 °K. Die punktierte Kurve zeigt die Absorption nach einer Temperung auf 100 °C. Untere Reihe: entsprechende Messungen für KI+10 At. % KF.

die bei verschiedenen Aufdampftemperaturen T_0 kondensiert worden sind. Die punktierte Kurve gibt zum Vergleich jeweils die ebenfalls bei 20 °K gemessenen Spektren der auf 100 °C getemperten störungsfreien Schichten wieder. Mit abnehmender Kondensationstemperatur erscheinen die Absorptionskurven immer stärker beeinflußt. Die schmale Bande der ungestörten Schicht spaltet auf in eine stark verbreiterte langwellige Komponente und eine ein wenig kurzwellig versetzte Bande. Fischer deutet die kurzwellige Bande als die Absorption sehr kleiner, durch die Oberflächenspannung kontrahierter Kristallite, den langwelligen Anteil als die Absorption stark fehlgeordneter aufgelockerter Bereiche (vgl. das Modellbild Fig. 2). Das untere Teilbild der Fig. 8 zeigt entsprechende Messungen an KI-Schichten, die durch gleichzeitige Kondensation von 10% KF noch stärker gestört worden sind. Das KF besitzt

selber in dem dargestellten Wellenlängenbereich keine Absorption. Man findet bei den tiefsten Kondensationstemperaturen nur noch die langwellig verschobene Bande. Die Schicht enthält keine geordneten Kristallite mehr (vgl. das Modellbild Fig. 7). Bei höheren Aufdampftemperaturen zeigt sich im Vergleich zum oberen Teilbild deutlich die Stabilisierung der Gitterstörungen durch den Zusatz an KF. Durch Tempern auf 100 °C läßt sich das KF zur Ausscheidung bringen. Man erhält wieder die (punktierte) Absorption des ungestörten KI.

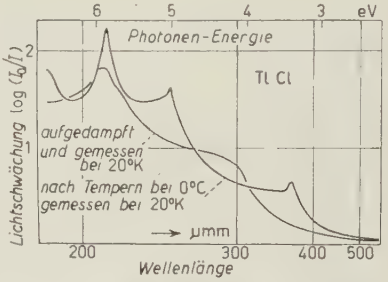


Fig. 9. Absorptionsspektrum einer bei 20 °K aufgedampften TlCl-Schicht: dick ausgezogene Kurve sofort nach der Kondensation; dünne Kurve nach Temperung bei 0 °C, gemessen bei 20 °K.

Als zweites Beispiel zeigt Fig. 9 eine Messung von ZINNGREBE [15] über den Einfluß der Fehlordnung auf die Absorption von TlCl. An Stelle der scharfen Absorption des ungestörten TlCl erhält man in der bei 20 °K kondensierten amorphen Schicht nur breite, verwaschene Banden. Interessant ist, daß die langwelligste Bande in diesem Fall im amorphen Zustand deutlich kurzwellig verschoben liegt, im Gegensatz zu dem oben betrachteten Verhalten des KI.

Bisher haben wir den Einfluß der beim Aufdampfen eingefrorenen Fehlordnung auf die optische Absorption betrachtet. In Fig. 10 vergleichen wir diesen mit dem Einfluß der durch die termischen Gitterschwingungen bedingten Unordnung (MARTIENSSEN [16]). Das obere Teilbild zeigt das Spektrum einer getemperten KI-Schicht bei verschiedenen Temperaturen, das untere Teilbild wiederholt noch einmal das Spektrum einer amorphen und einer auskristallisierten Schicht bei tiefen Temperaturen. Man erkennt deutlich die überraschende Ähnlichkeit zwischen der Absorption der auf 480 °K erhitzten kristallinen Schicht mit der der amorphen Schicht bei tiefen Temperaturen. Temperaturbewegung und ein-

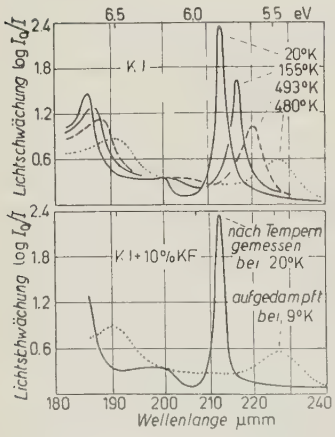


Fig. 10. – Oberes Teilbild: Einfluß der Meßtemperatur auf die Absorption einer getemperten KI-Schicht. Unteres Teilbild: Einfluß der durch abschreckende Kondensation bei 9 °K eingefrorenen Fehlordnung auf die Absorption einer Schicht aus KI + 10 At. % KF. Der KF-Zusatz dient zur Behinderung der Kristallisation des KI. Die Absorption von KF liegt außerhalb des dargestellten Wellenänlgenbereiches.

gefrorene Gitterstörungen wirken also in gleicher Weise auf die Lage und Gestalt der Excitonenbande ein.

Es liegt nahe, die lineare Verschiebung des Bandenmaximums bei zunehmender Temperatur mit der linearen thermischen Ausdehnung des Kristallgitters in Zusammenhang zu bringen. Die Breite der Absorptionsbande läßt sich dann mit den durch die thermischen Gitterschwingungen bedingten Dichteschwankungen des Gitters verstehen. Je nach der momentanen Konfiguration der direkten Gitterumgebung absorbiert ein einzelnes Ionenpaar einmal etwas kurzwelliger, einmal etwas langwelliger. Die gemessene breite Absorptionsbande bei hohen Temperaturen ist das zeitliche und räumliche Mittel über diese mit den Gitterschwingungen fluktuierende Verteilung. Im Rahmen dieses Modells können wir aus der gemessenen Halbwertsbreite leicht die Amplitude der Gitterschwingungen berechnen. Aus dem Verhältnis der Bandenverschiebung mit der Temperatur und dem thermischen Ausdehnungskoeffizienten

$$\frac{\partial h\nu_{\max}}{\partial T} \cdot \frac{1}{l} \frac{\partial l}{\partial T} = 21 \text{ eV},$$

folgt eine Bandenverschiebung um 21 eV für eine relative Längenänderung um 100%. Die Halbwertsbreite der Bande beträgt z.B. bei Zimmertemperatur ca. 0.2 eV, entspricht also einer Schwankung der Längenkoordinate um ca. 1%. Wir kommen also zu dem Ergebnis, daß die mittlere Amplitude der Wärmeschwingungen bei Zimmertemperatur in der Größenordnung 1% der Gitterkonstanten liegt, in Übereinstimmung mit den aus Röntgenbeugungsaufnahmen ermittelten Werten.

Fig. 11 zeigt eine ausführliche Messung der Temperaturabhängigkeit des langwelligen Ausläufers der Excitonenbande in KBr, dargestellt in einem logarithmischen Ordinatenmaßstab. Der Verlauf der Absorptionskonstante K läßt sich in Abhängigkeit von der Frequenz ν über einen Bereich von 6 Zehnerpotenzen und für Temperaturen bis zu 1000 °K mit der Gleichung beschreiben

$$K = K_0 \exp\left[-\sigma \frac{h\nu_{\max} - h\nu}{kT}\right]$$

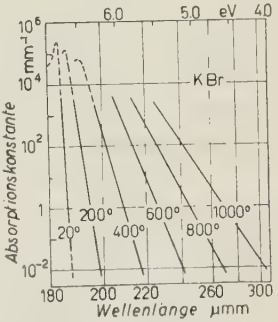


Fig. 11. – Verlauf der optischen Absorptionskante von KBr bei sechs verschiedenen Temperaturen zwischen 20 °K und 1000 °K.

(vgl. URBACH [17]). Darin haben die beiden Konstanten K_0 und σ für KBr die Werte $K_0 = 2.4 \cdot 10^5\ mm^{-1}$ und $\sigma = 0.79$; $\nu_{\max}$ ist die Frequenz des Bandenmaximums bei der Temperatur T. Nur bei tiefen Temperaturen, unterhalb 150 °K, findet man Abweichungen von dieser Beziehung. Die Gleichung liefert

bei verschwindender Temperatur eine unbegrenzt steil abfallende Absorptionsbande, gemessen wird aber bei den tiefsten Temperaturen (20 °K) eine Steilheit, die nach der Gleichung einer Temperatur von 69 °K entspricht. Wenn man die Breite der Absorptionsbande mit den durch die thermischen Gitterschwingungen bedingten Dichteschwankungen im Kristall erklärt, so ist die endliche Breite der Absorptionsbande bei verschwindender Temperatur ein Maß für die Dichteschwankungen beim absoluten Nullpunkt oder für die Nullpunktenergie des Kristalls. Man findet auf diese Weise eine Nullpunktsenergie von 0.035 eV für jedes Molekül, in guter Übereinstimmung mit dem aus Messungen der spezifischen Wärme von CLUSIUS und Mitarbeitern [18] folgenden Wert von 0.030 eV.

Nachdem wir den Einfluß der Temperatur auf die Gestalt der Excitonenbande kennengelernt haben, kommen wir zurück auf den in Fig. 10 ausgedrückten Vergleich zwischen der Beeinflussung der Absorption durch die Temperaturbewegung und durch eingefrorene Gitterstörungen. Lage und Breite der Bande bei hohen Temperaturen haben wir in unserem Modell durch die mittlere Größe und mittlere Schwankung des Ionenabstandes im Kristallgitter erklärt. Entsprechend deuten wir die verschobene und verbreiterte Absorption in der amorphen Schicht durch den aufgeweiteten und räumlich schwankenden Ionenabstand in dieser fehlgeordneten Schicht. Die in der Schicht enthaltene (potentielle) Fehlordnungsenergie sollte dann gleich der Hälfte der (potentiellen + kinetischen) thermischen Energie in der kristallinen Schicht bei 480 °K sein. Ist C_v die spezifische Wärme und N die spezifische Molekülzahl, so ergibt sich daraus für jedes Molekül der gestörten Schicht eine mittlere Fehlordnungsenergie

$$\varepsilon = \frac{1}{2N}\int_0^{480\,^\circ\mathrm{K}} C_v \,\mathrm{d}T = 0.11 \text{ eV},$$

das ist etwas mehr, als die halbe Schmelzwärme von 0.18 eV. Wie in Abschnitt 5 gezeigt wird, führen kalorimetrische Messungen der Fehlordnungsenergie amorpher Metallschichten zu ganz entsprechenden Werten.

Eine Übereinstimmung der optischen Eigenschaften stark gestörter Kristalle mit denen von ungestörten Kristallen bei hohen Temperaturen ist schon oft diskutiert worden (R. W. POHL [19] 1928). Interessant ist hier vielleicht noch das Beispiel des TlCl. In Fig. 9 haben wir gesehen, daß TlCl im amorphen Zustand eine kurzwellig verschobene Absorptionskante besitzt, im Gegensatz zu fast allen anderen Salzen, bei denen die Absorptionskante durch Gitterstörungen langwellig verschoben wird. Die Tl-Salze stellen aber auch hinsichtlich der Temperaturverschiebung der langwelligsten Absorptionsbande eine Ausnahme dar. In diesen Salzen verschiebt sich die Bande — wie PLEIL [20] zuerst ausführlich gezeigt hat — mit zunehmender Temperatur zu kurzen

Wellen. Auch hier bleibt also, jedenfalls qualitativ, der übereinstimmende Einfluß von Temperatur und Gitterstörungen erhalten.

4. – Elektrischer Widerstand und Supraleitung.

Während die optische Absorption besonders bei den Salzkristallen Aufschluß über die Gitterfehlordnung zu geben vermag, da wir es hier mit schmalen voneinander getrennten Absorptionsbanden zu tun haben, ist für die Untersuchung des Ordnungszustandes von metallischen Schichten die Messung ihrer elektrischen Leitfähigkeit besonders geeignet. Dabei zeigt sich bei sehr tiefen Temperaturen, daß auch die Supraleitfähigkeit der Metalle entscheidend beeinflußt werden kann durch den Grad der Fehlordnung.

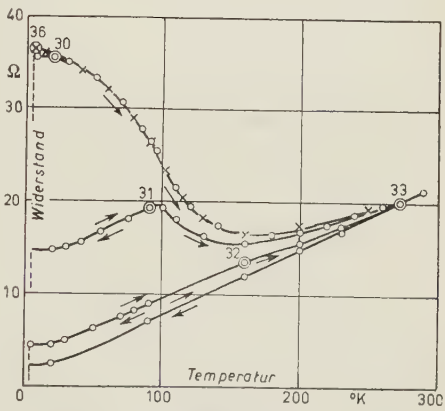


Fig. 12. – Widerstandsverhalten von Sn-Schichten mit verschiedenen Kondensationstemperaturen. Die Kondensationstemperaturen der durch Zahlen gekennzeichneten Schichten sind durch einen Doppelkreis markiert. Schichtdicke 500 Å.

Fig. 12 zeigt zunächst am Beispiel einer Aufdampfschicht aus reinem Zinn den Einfluß der Kondensationstemperatur (gekennzeichnet durch einen Doppelkreis ◎) auf den elektrischen Widerstand (Buckel und Hilsch[21]). Eine bei Zimmertemperatur erzeugte Schicht ergibt beim Abkühlen die normale lineare Abnahme des Widerstandes mit der Temperatur. Bei tiefen Temperaturen verbleibt ein Restwiderstand, der erheblich verringert werden kann, indem die Schicht längere Zeit in Schmelzpunktnähe getempert wird. Dieser Restwiderstand zeigt also eine gewisse Störung bereits der bei Zimmertemperatur kondensierten Schicht an. Je tiefer die Kondensationstemperatur T_0 gewählt wird, umso höher liegt der elektrische Widerstand. Alle Schichten ergeben unterhalb der Kondensationstemperatur einen linearen, reversiblen Verlauf des Widerstandes, oberhalb aber eine irreversible Abnahme infolge einer Ausscheidung der Gitterstörungen. Der bei tiefen Temperaturen verbleibende Restwiderstand kann als Maß für den Grad der Fehlordnung dienen. Aus dem parallelen Verlauf der reversiblen Kurvenabschnitte unterhalb T_0 folgt die Gültigkeit der Mathiessen'schen Regel auch für die sehr stark gestörten Schichten. Der temperaturabhängige Anteil des spezifischen Widerstandes stimmt in allen Fällen mit dem spezifischen Widerstand von kompaktem Zinn überein (Niebuhr [22]).

Besonders interessant ist das Widerstandsverhalten der Zinn-Schichten bei sehr tiefen Temperaturen (Fig. 13). Alle fünf (in Fig. 12 und 13 durch Zahlen gekennzeichneten) Schichten werden supraleitend, haben aber ganz verschiedene Sprungtemperaturen. Nur die bei 290 °K kondensierte Schicht zeigt eine Sprungtemperatur, die nahezu mit der des kompakten Zinns (3.7 °K) übereinstimmt. Alle anderen Schichten werden bereits bei höheren Temperaturen supraleitend, und zwar ist die Sprungtemperatur umso höher, je tiefer die Kondensationstemperatur, d.h. je stärker die Fehlordnung ist. Die Schichten stärkster und schwächster Störung weisen sehr scharfe Übergangskurven auf, bei den übrigen deutet der breite Abfall der Sprungkurven auf Inhomogenitäten im kristallinen Aufbau dieser Schichten hin.

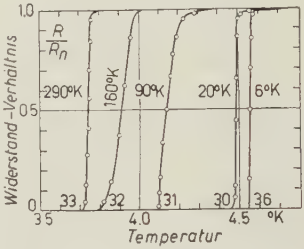


Fig. 13. – Übergangskurven zur Supraleitung der gleichen bei verschiedenen Temperaturen kondensierten Zinn-Schichten wie in Fig. 12.

Ähnliche Ergebnisse haben BUCKEL und HILSCH [23] auch an einer ganzen Reihe anderer Metalle erhalten: Al, Zn, In, Tl, Pb und Hg. Die Widerstandskurven der abgeschreckt kondensierten Schichten unterscheiden sich in der Lage des Minimums beim Tempern der Schichten (vgl. Fig. 12). Bei der Temperatur des Minimums ist der irreversible Abbau der Gitterstörungen so weit fortgeschritten, daß der thermische Widerstandsanstieg der Metallschicht zu überwiegen beginnt. Dieses Minimum liegt bei allen genannten Metallen etwa bei der jeweiligen Debye-Temperatur Θ, bei Aluminium am höchsten, mit über 300 °K und bei Hg am tiefsten mit ca. 70 °K. Bei allen Metallen ist also in gleicher Weise bei der Temperatur Θ der überwiegende Anteil der Gitterstörungen ausgeschieden.

Man wird erwarten, daß die Metalle mit hohen Θ bei fester Kondensationstemperatur von 4 °K auch die höchste Fehlordnung und nach dem Ergebnis der Fig. 13 dann auch die stärkste Erhöhung ihrer Sprungtemperatur zeigen. Tatsächlich erkennt man in Fig. 14 einen angenähert linearen Zusammenhang zwischen der Debye-Temperatur und der Sprungpunktserhöhung durch

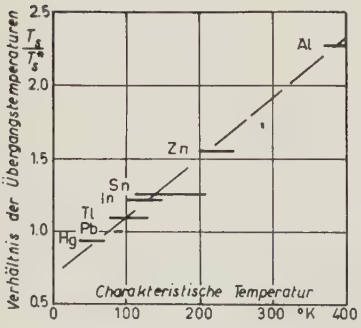


Fig. 14. – Verhältnis der Übergangstemperaturen T_s/T_s^* in Abhängigkeit von der charakteristischen Debye-Temperatur. T_s Übergangstemperatur der bei 4 °K kondensierten Schicht sofort nach der Entstehung; T_s^* Übergangstemperatur des kompakten Materials. Die horizontalen Striche bezeichnen den Bereich der Literaturwerte für die charakteristische Temperatur des angegebenen Metalls.

abschreckende Kondensation. Im Falle des Al wird der Sprungpunkt um den Faktor 2.27 erhöht, im entgegengesetzten Fall des Hg wird sogar eine geringe Erniedrigung der Sprungtemperatur beobachtet.

Ein besonderes Verhalten zeigen wieder Wismut und Gallium, die sich schon in Abschnitt 2 durch das Auftreten einer amorphen Phase von den übrigen Metallen unterschieden hatten. Kompaktes Wismut wird bis 0.05 °K herab nicht supraleitend. Dampft man es aber bei 4 °K auf, so erhält man eine supraleitende Schicht (Hilsch [24]) mit scharfer Übergangstemperatur bei 6 °K (Fig. 15). Das Wismut ist in diesem Zustand amorph. Beim Erwärmen tritt bei 15 °K eine starke irreversible Widerstandszunahme auf. Bei dieser Temperatur zeigen die Beugungsaufnahmen den Übergang vom amorphen in den kristallinen Zustand. Bei erneutem Abkühlen weist die Schicht das normale Verhalten des Bi-Metalles auf, die Supraleitfähigkeit ist verschwunden.

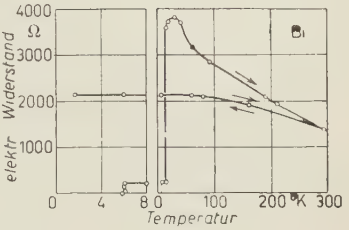


Fig. 15. – Widerstandsverhalten einer bei 2 °K kondensierten Wismut-Schicht. Schichtdicke ca. 500 Å. ⇄ reversibler Widerstandsverlauf, → irreversibler Widerstandsverlauf.

Die Zunahme des elektrischen Widerstandes beim Tempern steht im Gegensatz zu allen bisher diskutierten Beispielen. Wismut und Gallium zeichnen sich vor den meisten anderen Metallen dadurch aus, daß sie unter Volumen- und Widerstandsabnahme schmelzen. Bei der Kondensation bei tiefen Temperaturen entsteht das Metall offenbar ebenfalls in einer dichteren und besser leitenden Form als die des normalen Bi-Gitters. Bereits die Beugungsaufnahmen hatten ja nach dem Aufdampfen einen der Schmelze ähnlichen Zustand gezeigt. Daß Wismut in dieser dichter gepackten Modifikation supraleitend wird, steht im Einklang mit dem Befund von Chester und Jones [25], daß kompaktes Bi unter allseitigem Druck von mindestend $2 \cdot 10^4$ atü bei etwa 6 °K supraleitend wird.

Ein wenig komplizierter ist das Verhalten aufgedampfter Gallium-Schichten. Die Sprungtemperatur von normalem Gallium-Metall liegt bei 1.7 °K, eine bei 4 °K aufgedampfte Ga-Schicht bleibt dagegen bis herauf zu 8.4 °K supraleitend. Hier ist die Sprungtemperatur also nahezu um den Faktor 8 erhöht worden. Beim Erwärmen tritt bei 14 °K eine erste steile Widerstandsabnahme auf — gleichzeitig mit der Umwandlung der amorphen in die bisher unbekannte kristalline Phase. Bei erneutem Abkühlen findet man eine Sprungtemperatur von 6.3 °K. Erwärmt man die Schicht über 60 °K hinaus, so erfolgt eine zweite irreversible Widerstandsänderung, jetzt aber eine Widerstandszunahme, ähnlich wie beim Bi. Nach der zweiten Umwandlung besitzt das Ga wieder seine normale Sprungtemperatur von 1.07 °K. Nach dem

Widerstandsverhalten und den Beugungsaufnahmen müssen wir die erste Umwandlung mit der Bildung eines geordneten Gitters, die zweite Umwandlung mit dem Übergang aus einem dicht gepackten Gitter zu dem weniger dichten normalen Ga-Gitter deuten.

Eine ähnlich starke Veränderung der Sprungtemperatur, wie sie Bi und Ga im amorphen Zustand zeigen, kann auch in abgeschreckt kondensierten Sn-Schichten hervorgerufen werden, wenn die Fehlordnung durch gleichzeitige Kondensation von Fremdbausteinen noch erhöht wird (HILSCH und SCHIERTEL [26]). Eine bei 4 °K kondensierte Sn+10 At. % Cu-Schicht (vgl. Fig. 6) hat eine Sprungtemperatur von 7 °K (gegenüber 4.6 bzw. 3.7 °K von reinem, abgeschreckt kondensierten bzw. kompaktem Sn). Der Restwiderstand liegt sehr hoch, zeigt aber bereits beim Erwärmen auf ca. 20 °K einen steilen, irreversiblen Abfall. In diesem Temperaturbereich beginnt die Schicht nach Fig. 6 zu kristallisieren. Ein zweiter Abfall bei 230 °K deutet die Bildung der intermetallischen Verbindung Sn-Cu an. Bei erneutem Abkühlen weist die Schicht wieder die normale Sprungtemperatur von 3.7 °K auf.

Die starke Erhöhung der Sprungtemperatur auf 7 °K beruht keineswegs auf einer spezifischen Wirkung des Kupfers: sie läßt sich ebenso gut durch viele andere Metalle hervorrufen. Untersucht wurden Zusätze von Co, Ni, Zn und As. Neuerdings konnte BARTH [27] zeigen, daß auch nichtmetallische Zusätze, z.B. SiO die Sprungtemperatur in gleicher Weise beeinflussen. Die Wirksamkeit des Zusatzes liegt offenbar lediglich in der Behinderung der Kristallisation des Sn. Das geht besonders deutlich aus einer Untersuchung der Abhängigkeit der Sprungtemperatur von der Konzentration der verschiedenen Zusatzsubstanzen hervor [26]. Die Sprungtemperatur beginnt deutlich zu steigen bei Zusatzkonzentrationen von einigen Prozent, erreicht dann je nach Art des Zusatzes mehr oder weniger schnell ein Maximum und fällt schließlich mit weiter steigender Konzentration angenähert linear ab. In diesem letzten Bereich macht sich die spezifische Wirkung des Zusatzes in einer Erniedrigung der Sprungtemperatur bemerkbar. Extrapoliert man den linearen Abfall zu sehr kleinen Konzentrationen, so gelangt man unabhängig von der Art des Zusatzes auf eine Sprungtemperatur des Sn von ungefähr 8 °K. Würde es also gelingen, auch eine zusatzfreie Zinn-Schicht genügend fein zu zerteilen, so würde dieses «amorphe» Zinn eine Sprungtemperatur von ungefähr 8 °K besitzen.

Wie auch immer der Einfluß der Fehlordnung auf die Sprungtemperatur der metallischen Schichten gedeutet werden mag, in jedem Fall zeigen diese Experimente eindeutig, daß die Erscheinung der Supraleitung keineswegs an das Vorhandensein ungestörter «Idealkristalle» gebunden ist, sondern daß die Sprungtemperatur sogar in stark fehlgeordneten oder gar amorphen Metallen wesentlich höher liegen kann als in den entsprechenden Substanzen in kristalliner Form.

5. – Messung der Fehlordnungsenergie.

Zum Schluß sei noch kurz auf einige neuere Messungen von SANDER [28] zur Bestimmung der in den gestörten Schichten gespeicherten Fehlordnungsenergie eingegangen. Das benutzte Kalorimeter besteht aus einer 35 μm dicken Aluminium-Folie, die mit wenigen Nylon-Fäden thermisch isoliert in einem Kupferrahmen aufgehängt ist, der selber in Wärmekontakt mit der Kühlflüssigkeit (fl. H_2) steht. Gekühlte Abschirmbecher verhindern die Wärmezustrahlung von den Wänden des Vakuumgefäßes. Durch eine von außen zu bedienende Klammer kann die Aluminium-Folie mit dem Kupferrahmen in thermischen Kontakt gebracht werden. Nachdem die Folie mit der zu untersuchenden Schicht bedampft ist, wird sie durch Lösen der Klammer thermisch isoliert und mit Hilfe einer Bestrahlung mit konstantem Lichtstrom gleichmäßig aufgewärmt. Ein Thermoelement registriert laufend die Temperaturdifferenz zwischen der Folie und dem Kupferrahmen.

Fig. 16 zeigt das Ergebnis einer derartigen Messung am Beispiel einer Wismut-Schicht. Durch einen Zusatz von 15% Antimon wird der amorphe Zustand der aufgedampften Wismut-Schicht soweit stabilisiert (BARTH [29]), daß die Kristallisation statt bei 15 °K (vgl. Fig. 15) erst bei ca. 30 °K einsetzt. Die Aufwärmkurve mit einer frisch kondensierten Schicht zeigt eine deutliche Stufe, während eine Wiederholung der Messung nach erneutem Abkühlen, wie zu erwarten, einen gleichmäßigen Anstieg der Temperatur liefert. Aus der bekannten spezifischen Wärme und Masse der Al-Folie kann die bei der Umwandlung des Wismuts entwickelte Wärmeleistung berechnet werden (unteres Teilbild der Fig. 16). Insgesamt erhält man für jedes Bi-Atom eine mittlere Umwandlungswärme von 0.063 eV, das ist zum Vergleich etwas mehr als die Hälfte der Schmelzwärme des Bi von 0.12 eV. Bei Gallium findet SANDER bei den beiden in Abschnitt 2 und 4 diskutierten Umwandlungen eine Wärmetönung von $0.013 + 0.025 = 0.038$ eV für jedes Atom, das ist wieder etwas mehr als die Hälfte der Schmelzwärme von 0.058 eV. Diese Bestimmungen der Fehlordnungsenergie bestätigen also den aus den Beugungsaufnahmen gezogenen Schluß, daß in den abgeschreckt kondensierten Bi- und Ga-Schichten ein der Schmelze ähnlicher, amorpher Zustand vorliegt.

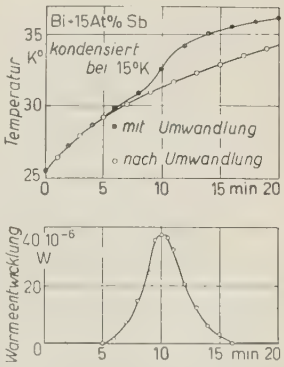


Fig. 16. – Zur Messung der Umwandlungswärme. Oberes Teilbild: Aufwärmung einer bei 15 °K kondensierten Schicht aus Bi+15 At. % Sb mit und nach Umwandlung aus dem flüssigkeitsähnlichen Zustand. Schichtdicke 2000 Å. Unteres Teilbild: berechnete freiwerdende Wärmeleistung.

Die bei der ersten Umwandlung der Ga-Schicht freiwerdende Wärmemenge (0.013 eV pro Atom) ist verhältnismäßig gering. Man könnte daraus schließen, daß die neue bei der Umwandlung entstehende Ga-Modifikation eine ähnliche Packung der Atome hat wie die im amorphen Zustand. Auch aus den Widerstandsmessungen folgte ja, daß das neue Gallium-Gitter einen dichter gepackten, der Schmelze ähnlicheren Zustand darstellt. Da bei der Umwandlung der neuen Modifikation in das normale Gallium-Gitter wieder Wärme frei wird, ist das neue Gitter als eine Hochtemperaturmodifikation des Ga anzusprechen. Die abschreckende Kondensation liefert also bei der tiefsten Temperatur einen der Schmelze ähnlichen Zustand, bei etwas höherer Temperatur einen kristallinen Zustand, der im Gleichgewicht nur bei höheren Temperaturen existieren kann und erst bei noch höherer Kondensationstemperatur das unter normalen Bedingungen stabile Gitter. Es ergibt sich also auch hier wieder die Folgerung, daß der bei der abschreckenden Kondensation erzielte metastabile Zustand der Aufdampfschichten mit einem Gleichgewichtszustand bei hohen Temperaturen vergleichbar ist, und zwar entspricht der bei der tiefsten Kondensationstemperatur erreichte Ordnungsgrad dem zu der höchsten Temperatur gehörigen Gleichgewichtszustand.

LITERATURVERZEICHNIS

[1] W. Rühl: *Zeits. f. Phys.*, **143**, 591 (1956).
[2] R. Hilsch: *Zeits. f. Phys.*, **138**, 432 (1954).
[3] M. Born und J. E. Mayer: *Zeits. f. Phys.*, **75**, 15 (1932).
[4] W. Rühl: *Zeits. f. Phys.*, **143**, 591 (1956).
[5] W. Buckel: *Zeits. f. Phys.*, **138**, 136 (1954).
[6] H. J. Queisser: *Dissertation*, Göttingen.
[7] R. Hilsch und R. W. Pohl: *Zeits. f. Phys.*, **59**, 812 (1930).
[8] E. G. Schneider und H. M. O'Bryan: *Phys. Rev.*, **51**, 293 (1937).
[9] J. Frenkel: *Phys. Rev.*, **37**, 17, 1276 (1931).
[10] G. H. Wannier: *Phys. Rev.*, **52**, 191 (1937).
[11] H. R. Phillipp und E. A. Taft: *Journ. Phys. Chem. Solids*, **1**, 159 (1956); sowie E. A. Taft und H. R. Philipp: *Journ. Phys. Chem. Solids*, **1**, 159 (1957). Herrn Dr. L. Apker sei herzlich gedankt für die vorzeitige Überlassung eines Manuskriptes der letztgenannten Arbeit.
[12] S. Tutihasi: *Phys. Rev.*, **105**, 882 (1957).
[13] Y. Okamoto: *Nachr. Akad. Wiss. Göttingen*, Nr. **14**, 275 (1956).
[14] F. Fischer: *Zeits. f. Phys.*, **139**, 328 (1954).
[15] H. Zinngrebe: *Dissertation*, Göttingen.
[16] W. Martienssen: *Journ. Phys. Chem. Solids*, **2**, 257 (1957).
[17] F. Urbach: *Phys. Rev.*, **102**, 1519 (1956).
[18] K. Clusius, J. Goldman und A. Z. Perlick: *Naturforsch.*, **4a**, 424 (1949).
[19] R. W. Pohl: *Naturwiss.*, **16**, 477 (1928).

[20] V. Pleil: *Diplomarbeit*, Erlangen 1953.
[21] W. Buckel und R. Hilsch: *Zeits. f. Phys.*, **132**, 420 (1952).
[22] J. Niebuhr: *Zeits. f. Phys.*, **132**, 468 (1952).
[23] W. Buckel und R. Hilsch: *Zeits. f. Phys.*, **138**, 109 (1954).
[24] R. Hilsch: *Phys. Verh.*, **2**, 101 (1951) (Erlanger Tagung).
[25] P. F. Chester und G. O. Jones: *Phil. Mag.*, **44**, 1284 (1953).
[26] R. Hilsch und A. Schertel: unveröffentlicht.
[27] N. Barth: unveröffentlicht.
[28] W. Sander: *Zeits. f. Phys.*, **147**, 361 (1957).
[29] N. Barth: *Zeits. f. Phys.*, **142**, 58 (1955).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Color Centers in Alkali Halides.

H. PICK

Technische Hochschule - Stuttgart

1. – Historical comments.

The beginning of research on imperfections in solids dates from 1896 when E. GOLDSTEIN [1] observed the coloration of polar solids, especially alkali halides, by high energy-radiation. Later on it was found by Z. GYULAI [2] that the same colors appear in crystals with stechiometric excess of the metal, e.g. Na in NaCl. The colors are due to absorption bands in the visible range of the spectrum. Two important conclusions were drawn from this knowledge.

1) Simple methods to produce these colorations. They will be discussed in Sect. 3.

2) The coloration by high energy radiation may be understood as a dissociation process of a small amount of alkali halide molecules, that means the production of atomically dispersed alkali. Besides the alkali, the same quantity of neutral halogen must then be present. The optical absorption of this excess halogen was found in 1949 [3] in the ultraviolet. The absorption bands were called V-bands. Similar spectra had been observed by E. MOLLWO in 1935 [4] in KBr and KI with stechiometric excess of bromine and iodine.

The investigation of color centers in crystals was highly stimulated by a number of important experimental methods. A few of them may be mentioned.

1) The production of single crystals from the melt [5].

2) The use of a wide range of the optical spectrum from 0.1—50 μm to measure absorption and emission spectra.

3) The combination of dark and photoconductivity measurements with optical observations.

4) The introduction of dielectric, magnetic and thermodynamic data.

5) The application of a wide temperature range, especially low temperatures.

The number of publications concerning color centers in solids is of the order of 1000. The most important review articles were given by R. W. POHL [6] in 1937 and by F. SEITZ in 1946 and 1954 [7].

During the past 10 to 15 years a remarkable fraction of research in Physics was done in the solid state field, mainly in semiconductor-work. Before this time, the main interest in solids was directed to the alkali halides, and till now the best understood point imperfection is the color-center in alkali halides. This center is often used as a model for many problems of solid state physics. This is the reason why three lectures of this Summer-School are devoted to the color-centers in alkali halide crystals.

2. – The most important imperfections in alkali halides.

Let us consider a single crystal of a carefully purified alkali halide, for instance KCl; then the concentration of impurities is less than 10^{-5} or 10^{-6}. Apart from these chemical imperfections there are the following structural imperfections (Fig. 1):

1) *Dislocations.* Dr. AMELINCKX gave a seminar on this topic.

2) *Vacancies.* Namely positive ion vacancies and negative ion vacancies (α-center). The concentration increases exponentially with temperature. These vacancies are responsible for ionic conductivity and diffusion processes at higher temperatures.

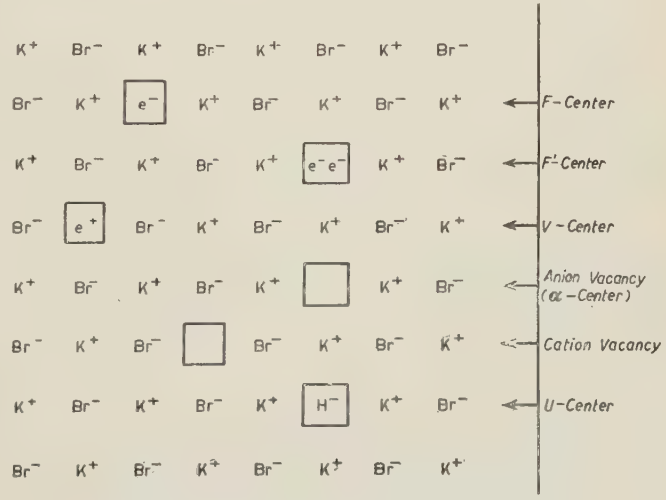


Fig. 1. – Point imperfections in alkali halides.

3) *Aggregates of vacancies*, as pairs and groups of 3, 4 and more vacancies. They are formed if the temperature of the crystal is decreased.

4) Possibly there are some ions in *interstitional positions.*
In addition to these imperfections there are imperfections which are occupied by electrons or holes.

5) Electrons may be trapped in halogen ion vacancies. This imperfection is called an *F-center.* Our interest will be directed to this F-center.

6) An F-center is able to bind another electron. The center so formed is called an *F'-center.*

7) The F and F'-center may be associated with other vacancies and pairs of vacancies. These centers have been named as M, N, R or A, B, C, D, E, G, centers.

8) Another group of imperfections is formed by the trapping of holes in vacancies and aggregates of vacancies. They have been named V-centers, as V_1, V_2, V_3, ...

9) It should be mentioned that there are good reasons for the assumption of a center which is represented by a « *self-trapped hole* ». Dr. KÄNZIG will give some details in his seminar. The self trapped electron has not yet been observed.

I wish to add two remarks:

1) There are only a few point imperfections the properties of which are well known. These are the F-center, the F'-center and the α-center; for all other centers it is not yet possible to combine experimental observations and physical properties with well proved atomic models. This is the situation for all the aggregate-centers, as M, R and N, and for all the V-centers.

2) There are many types of combinations of imperfections, but only those exist which preserve the electric neutrality of the crystal. For instance: excess electrons are permitted to enter the crystal if the same number of halogen ions diffuses out of the crystal.

3. – Methods for the production of F-centers.

3·1. – The technically simplest method of coloration is that of irradiation with high energy radiation, as X-rays, γ-rays, fast electrons.... A first order description of the mechanism of the photochemical coloration may be given as follows. By the absorption of high energy radiation many pairs of electrons and holes, and in addition vacancies and other structural imperfections are produced. A high fraction of the pairs of carriers will recombine with the emission of fluorescence light; a small fraction will be trapped. The electrons are trapped as F-centers (neutral alkali), the holes are trapped as V-centers

(neutral halogen). The amount of energy absorbed in the crystal, divided by the number of F-centers produced, is of the order of several 100 eV at the beginning of the coloration. It increases rapidly.

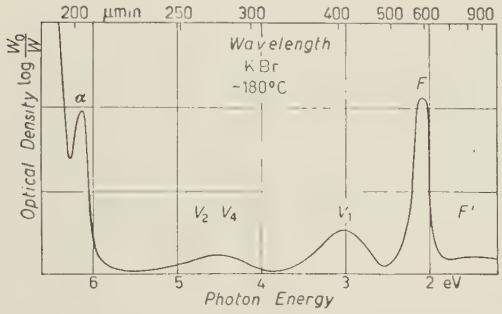


Fig. 2. - X-rayed crystal with absorption bands due to the presence of different types of centers.

This method has two disadvantages:

a) Not only F-centers but many other imperfections are produced at the same time. This is shown by the absorption spectrum of a crystal colored by radiation (Fig. 2).

b) The coloration is not stable. The crystal may easily be bleached by visible light or by heat.

3·2. – The photochemical coloration is much more stable if the crystal had been « sensitized ». A well understood example is KBr+KH [8]. In such a crystal there are negative H^--ions in positions of Br^--ions. The presence of these H^--ions is responsible for an ultraviolet absorption band, called U-band (Fig. 3). The U-band is reduced during X-irradiation and the absorption band of the F-centers is produced. There is no evidence for other imperfections. It may be concluded that the holes, produced during the absorption of high energy quanta are trapped by the negative H^--ions forming neutral H. The remaining electrons are trapped as F-centers. The efficiency of this process is much higher than that of process 1. At the beginning of the process the absorbed energy divided by the number of F-centers, which have been formed, is 40 eV.

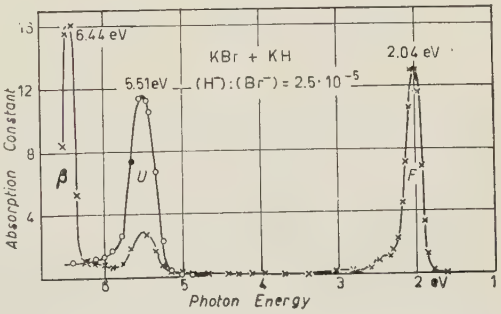


Fig. 3. – KBr+KH crystal before and after irradiation with X-rays.

3·3. – The method most easy to survey is the additive coloration of a crystal, that is the introduction of a stechiometric excess of alkali [9]. The process is as follows (Fig. 4).

A clear crystal and some pure alkali metal (for instance KCl and K) are sealed in a bomb of pyrex glass under good vacuum. The crystal is heated

up to a temperature of 700 °C and the metal is brought to a temperature of 300÷700 °C. In the course of several hours the metal diffuses from the vapour phase into the crystal. The equilibrium concentration inside the crystal is proportional to the vapour concentration and nearly equal to that of the vapour outside the crystal. It is given by the lowest temperature in the glass bomb. There is only a weak dependence on the temperature of the crystal. As a result one gets concentrations of F-centers of the order of $(10^{15} \div 10^{19})\text{cm}^{-3}$. By lowering the temperature of the crystal slowly down to room temperature the metal inside the crystal is associated in colloidal particles. By quenching the crystal, the high temperature state may be preserved: the excess metal is present in atomically dispersed form as F-centers. Such a crystal has only two new absorption bands: F and β. It was found that these bands are always combined (Fig. 3; the F-band at 2.04 eV; the β-band at 6.44 eV).

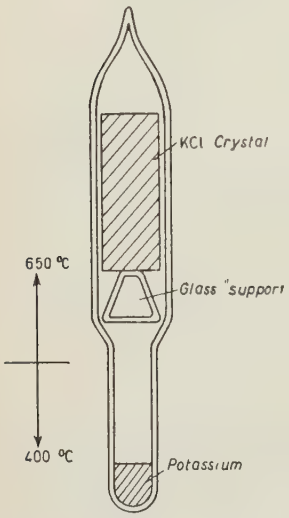


Fig. 4. – Glass bomb to color a crystal additively.

3 4. – A quite similar result may be obtained by electrolysis [10]. If ionic current is passed through a crystal as KI at an elevated temperature, potassium is deposited at the cathode. If the oxidation of potassium is prevented by a vacuum tight contact between electrode and crystal, the potassium diffuses into the crystal forming F-centers. These F-centers are drawn to the anode by the applied electric field. This effect of moving F-centers may be understood as thermal dissociation of the F-center into a negative ion vacancy and a free electron. The electron migrating to the anode is usually trapped by vacancies. The empty vacancies migrate to the cathode. They are replaced by vacancies newly formed at the anode. The production of negative ion vacancies as carriers at the anode means deposition of free neutral halogen as iodine. This neutral iodine should be free to evaporate otherwise it diffuses back into the crystal forming V- centers, which migrate to the cathode [11].

The following considerations concern F-centers; therefore we are interested mainly in the method of additive coloration, which gives the best conditions to avoid other imperfections.

4. – The nature of the F-center.

Prior to the discussion of the properties of the F-center let us consider the atomistic model. This will be helpful in the following sections.

The main ideas of this model were given by J. H. DE BOER in 1937 [12]. They find full justification and confirmation in all the available data. A careful discussion of the different models has been given by S. I. PEKAR in 1951 [13].

The simplest model of the F-center is that of a neutral alkali atom. It is derived from the presence of a stechiometric excess of metal, but it does not give any details of the microstructure. The more precise description given by DE BOER is the following: The electron is trapped in a negative ion vacancy, the negative charge of a halogen ion being replaced by the negative charge of an electron. Thus the electric neutrality of the crystal is preserved.

What is the state of this F-electron? The nearest neighbours are 6 K^+-ions. They produce a center of positive charge in the middle of the vacancy. Therefore in a first order treatment the F-center may be considered similar to a H-atom [14 *a*]. With a box-shaped potential distribution the energy terms would then be:

$$E_c = \frac{1}{8} \frac{h^2}{m^*} \frac{1}{d^2} k^2 ,$$

where k = quantum number,
h = Planck's constant,
m^* = effective mass of the electron,
d = lattice parameter.

The frequency of the optical transition from the ground state to the first excited state then is:

$$\nu_{12} = \frac{3}{8} \frac{h}{m^*} \frac{1}{d^2} .$$

This is in good agreement with experiment. Interpolation formulae for many alkali halides were given by E. MOLLWO in 1931 and by H. F. IVEY in 1947 [14 *b*]. A more substantial description may be given with the help of optical and electrical data. They shall now be given as the most important experimental results.

5. – The optical absorption.

Pure alkali halide single crystals are highly transparent in the wide range between the fundamental electronic absorption in the far ultraviolet (Fig. 5) and the absorption of the lattice vibration in the far infrared (Fig. 6); in KBr this means between $(0.2 \div 15)$ μm. If F-centers are present two absorption bands are always observed. The β-band just on the long wave side of the fundamental absorption, and the F-band normally situated in the visible part

of the spectrum. These bands are strongly associated with one another. The ratio of the absorption constants being 7/6. The β-band (193 μmm in KBr) has been interpreted as the fundamental electronic excitation of the crystal

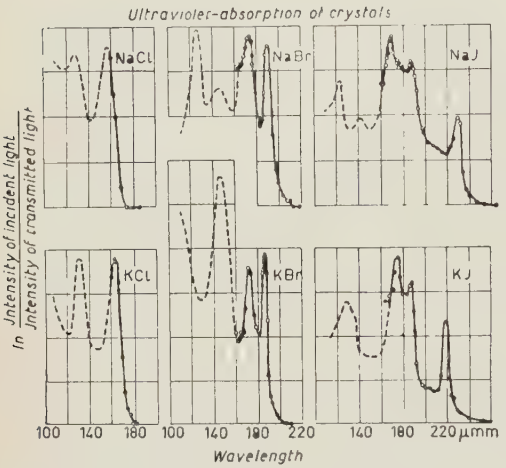


Fig. 5. – Fundamental absorption of several alkali halides.

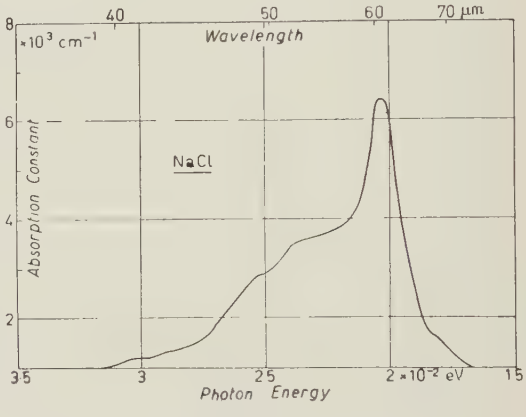


Fig. 6. – Infrared absorption of the lattice vibrations in NaCl.

in the presence of F-centers as imperfections. It is regarded as the first peak of the fundamental absorption shifted by the influence of the F-center. The height of the band then must be of the same order as the height of the F-absorption band. (A similar influence is observed when empty vacancies are present. They are responsible for the α-band). Our interest will mainly be directed to the F-band (610 μmm in KBr).

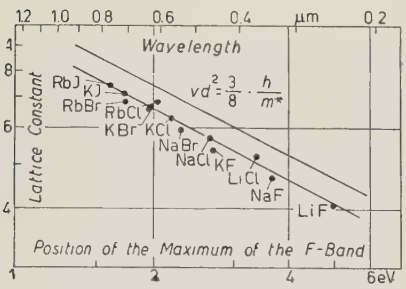


Fig. 7. – The dependence of the F-maximum on the lattice constant of the crystal.

The F-absorption is a bell shaped absorption curve. The frequency ν of the maximum is given by the lattice constant d. As mentioned in Sect. 4 one observes in good agreement with experience

$$\nu d^2 = \text{constant} .$$

This relation is good for pure crystals and mixed crystals of alkali halides (Fig. 7). One realizes that the properties of the F-center are mainly given by the lattice. They are proper lattice imperfections.

Table I summarizes the frequency and the half width of the F-band for different alkali halide crystals at room temperature. Both these values are strongly dependent on temperature as may be seen in Fig. 8 for KBr [15].

The absorption constant A of the maximum of the F-band and the half width H may be used to calculate the concentration N of absorbing electrons [16]. The classical theory of dispersion in a medium with refractive index n gives for the concentration of F-centers (A in cm^{-1}; H in eV)

$$Nf = \frac{18mn}{\pi e^2 h(n^2+2)^2} AH = \\ = 1.06 \cdot 10^{17} AH/\mathrm{cm}^3 \quad (\mathrm{KCl}),$$

with m, e, h, the usual atomic constants and f the oscillator strength. The value of f is nearly 1 (in KCl 0.81) [16a].

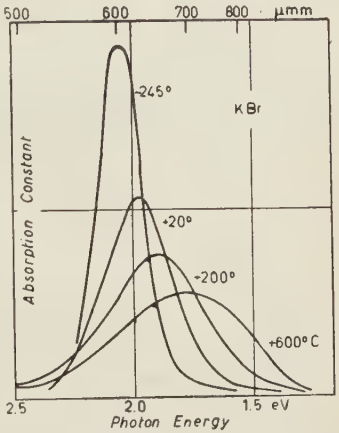


Fig. 8. – Temperature dependence of the F-band in KBr.

It should be mentioned that the half width of the F-band keeps a value of the order of 0.2 eV down to the lowest temperatures.

TABLE I.

	F	Cl	Br	J	
Na	λ 340 H 0.62	465 0.47	540 —	588 —	μmm eV
K	λ 455 H 0.41	563 0.36	630 0.365	685 0.35	μmm eV
Rb	λ H	624 0.31	720 0.28	775 —	μmm eV

λ Wavelength of the maximum of the F-band;
H Half-width of the F-band. 20 °C.

6. – The model of Pekar.

To get a more detailed knowledge on the structure of the F-center, the interaction between the electron of the imperfection and the surrounding lattice has to be taken into account. Otherwise one will never find the broad-

ness of the absorption band, the temperature dependence and the Stokes-shift. One of the best and for the practical application most useful theories was given by S. I. PEKAR [13] in 1949. PEKAR looks at the crystal as a dielectric medium with dielectric constant ε, refractive index n and frequency of the transversal optical vibrations ω_t (infrared absorption). The medium is polarized by the electron of the F-center and there is an interaction between the energy of the electron and the vibrational energy of the crystalline medium.

According to PEKAR the Hamiltonian has the following terms:

$$H = -\frac{\hbar}{2\mu}\Delta - \frac{e^2}{\varepsilon r} + e\int \frac{P(\boldsymbol{r})(\boldsymbol{r}-\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|^3}\,\mathrm{d}\tau_i + \sum \frac{2\pi}{c_w}\left(P_w^2 + \dot{P}_w^2\,\frac{1}{\omega_w^2}\right),$$

$-\frac{\hbar}{2\mu}\Delta$ — is the kinetic energy of the electron. The presence of the periodic potential of the lattice is considered by the introduction of the effective mass μ of the electron;

$-\frac{e^2}{\varepsilon r}$ — is the Coulomb-potential of the anion-vacancy, which is treated as positive point charge;

$e\int \frac{P(\boldsymbol{r}_i)(\boldsymbol{r}-\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}_i|^3}\,\mathrm{d}\tau_i$ — gives the polarization of the medium by the electron. Regarding this « adiabatic » approximation the movement of the electron is considered to be fast compared with the movement of the lattice polarization (displacement of the ions);

$P(\boldsymbol{r})$ — the polarizability is dipole moment/volume;

$\boldsymbol{r}_i$ — is the co-ordinate of the dielectric medium;

$\boldsymbol{r}$ — is the co-ordinate of the electron;

$\sum \frac{2\pi}{c_w}(\ldots)$ — may be considered as potential and kinetic energy of all lattice vibrations, that is the total spectrum of phonons.

But our interest is directed only to the interaction of the electron with the lattice, which is predominantly given by Coulomb forces. So we may disregard the acoustical modes and look only to the optical modes of the lattice vibration. But again this is more than we need. As opposed to the electromagnetic field which interacts only with the transversal optical vibrations the movement of an electron interacts only with the longitudinal modes. Looking at this spectrum of the longitudinal modes of the lattice vibration one realizes that only the frequencies ω_w with $w \to 0$ are of interest for our problem. The polarization given by the electron and any change of the polarization given by a change of the distribution of the electron extend over a range much more than one lattice constant d. So the waves of polarization excited by a moving electron

are long compared with the lattice constant

$$\frac{1}{w} \quad d\,.$$

For all these waves the dispersion may be neglected, they all have nearly the same frequency: $\omega_{\parallel}$; ω is combined with the corresponding transversal optical frequency $\omega_{\perp}$ by

$$\omega_{\parallel} = \omega_{\perp} \sqrt{\frac{\varepsilon}{n^2}}\,.$$

We end up with the following sum

$$\frac{2\pi}{c}\left(P^2 + \frac{1}{\omega_{\perp}} \dot{P}^2\right); \qquad c = \left(\frac{1}{n^2} - \frac{1}{\varepsilon}\right).$$

So far the discussion of the Hamiltonian, which takes into account the movement of the ions of the lattice by regarding the polarization of the dielectric medium, which is represented by the dielectric constant ε, the refractive index n and the infrared frequency $\omega_{\perp}$.

The problem has been solved by a variational method. The wave function of the system—electron plus vibrating lattice—is given by a product

$$\Psi = \psi \prod \Phi\,,$$

ψ being the electronic wave function,

$\prod \Phi$ being the product of the lattice vibrations considered.

The total energy then is given by the energy of the electron and the energy of the vibrating medium separately

$$H[\psi] = J[\psi] + \sum \hbar\omega_{\parallel}\left(n_w + \frac{1}{2}\right);$$

with

$$J[\psi] = \frac{h^2}{2\mu}\int |\operatorname{grad} \psi|^2 \,\mathrm{d}\tau - \frac{e^2}{\varepsilon}\int \frac{|\psi|^2}{r}\,\mathrm{d}\tau - \int PD\,\mathrm{d}\tau + \frac{2\pi}{c}\int\left(P^2 + \frac{1}{\omega_{\perp}^2}\dot{P}^2\right)\mathrm{d}\tau\,.$$

For the ground state of the electron a s-function is found to be well adapted

$$\psi_{1s} = A(1 - \alpha r)\exp\left[-\alpha r\right],$$

with

$$\alpha = \frac{\mu e^2}{2^{1/2}}\left(\frac{3}{\varepsilon} + c\right), \qquad A = \frac{\alpha^{\frac{3}{2}}}{\sqrt{7\pi}}\,.$$

For the first excited state of the electron the following p-function is found to be good—the lattice polarization remaining unchanged

$$\psi_{2p} = \frac{8}{\sqrt{2\pi}} (\alpha\beta)^{\frac{3}{2}} (\alpha\beta r) \exp[-2\alpha\beta r] \cos\vartheta ,$$

with

$$\beta = \frac{1 + 0.394\varepsilon r}{3 + \varepsilon e} .$$

The energy values of the ground state and the first excited state are the following

$$J[\psi_{1S}] + \hbar\omega_{\parallel} \sum_{w} (n_w + \tfrac{1}{2}) ,$$

$$J[\psi_{2P}] + \hbar\omega_{\parallel} \sum_{w} (n'_w + \tfrac{1}{2}) .$$

The energy level diagram is given by two electronic states superimposed by the terms of the lattice vibrations with constant energetic distance of the levels. $\frac{1}{2}\hbar\omega_{\parallel}$ is zero point energy of the longitudinal optical branch.

The absorption of the F-center now may be calculated as transition from the ground state to the first excited state with the help of the transition probabilities, which is the product of the constant electronic transition probability and the probability of the vibrational terms which change nearly steadily. The result is a number of lines, the envelope of which corresponds to the F-band. The number of the excited states of the vibration depends on temperature. Thus the temperature dependence of the maximum of the F-band

TABLE II.

	NaCl	KCl	RbCl	NaBr	KBr	RbBr	NaI	KI	RbI	
ε	5.8	4.78	5.2	6.39	4.81	5.16	6.6	5.2	5.58	
n^2	2.33	2.18	2.18	2.6	2.36	2.36	2.91	2.65	2.6	
$(h\nu)_{\max}$	2.65	2.19	1.98	2.29	1.96	1.71	2.1	1.8	1.59	eV
μ/m	2.78	1.85	1.78	2.96	1.87	1.70	3.25	2.11	1.89	
$\hbar\omega_{\parallel}$	0.032	0.026			0.020			0.017		eV
H_{calc}	0.47	0.40			0.37			0.35		eV
H_{obs}	0.47	0.36			0.365			0.35		eV

ε = dielectric constant;
n = refractive index;
$(h\nu)_{\max}$ = quantum energy of the maximum of the F-band;
μ/m = effective mass/mass of the free electron.
$\hbar\omega_{\parallel}$ = quantum energy of the infrared absorption maximum;
H = half width of the F-band.

and the half width may be deduced. All we need for the calculation is one experimental result—may be the F-band maximum at very low temperatures—to adjust for the effective mass μ and the constants of the material ε, n, ω . The temperature dependence of the frequency of the maximum, that of the half width of the band, the Stokes-shift of the emission may be calculated. As may be seen from Table II the experimental results are in rather good agreement with theoretical expectation. So the theory of PEKAR may be considered as a good first order solution of the problem. It is not able to find out the ratio of radiative and radiationless transitions. This can be done if the movement of the single ions surrounding the vacancy is considered.

7. – The excited center F^*.

As reported in Sect. **5** and **6** the transition of the F-electron from the ground state to the first excited state is the predominant transition. It has the highest probability. This excited state will now be discussed.

During the optical transition the microstructure of the center is that of the ground state (Franck-Condon principle). The excited electron influences the ions surrounding the vacancy; they are polarized and rearranged. The coupling of these ions with the total lattice leads to the excitation of lattice vibrations. So the electron may transfer energy to the lattice and vice versa the lattice to the electron.

At low temperatures the probability for the increase of the energy of the electron at the cost of the vibrational energy of the lattice is very small. The structure of the imperfection will be adjusted to the character of the excited electron. The state of polarization of the imperfection then will be given by the symmetry of the wave function of the electron. For a p-type-electron this would mean an axial microstructure of the excited F-center. During the displacement of the ions lattice vibrations are excited, phonons emitted and the crystal is heated (Sect. **11**).

The free energy of the total system being minimized the transition to the ground state may occur with the emission of fluorescent light (Sect. **10**). Owing to the energy transfer from the excited electron to the lattice the emission band is shifted to the long wave side compared with the absorption band.

If energy is transferred from the lattice to the electron, mainly in the high temperature range, the excited F-center may be ionized. The electron becomes mobile. One observes photoconductivity (Sect. **9**).

After some time the free electron will be retrapped. It was found that the ionized F-centers, the empty vacancies, as well as normal F-centers are effective traps for mobile electrons. If electrons are trapped by vacancies, F-centers are reformed. If electrons are trapped by F-centers, F'-centers are

produced (Sect. 8). These F'-centers are vacancies with two electrons in it. The ionization of the F'-center may be caused by optical excitation or by heat.

At low temperature the process of retrapping electrons as F-centers is combined with emission of light. At higher temperatures this luminescence is quenched and replaced by radiationless transitions (Sect. **11**).

8. – The F'-center.

Absorption of light in the spectral range of the F-band at low temperatures (in KCl below — 80 °C) reduces the F-band and produces a new band F' [17]. Under favourable conditions 80% of the electrons may be transferred from the F-state to the F'-state. Fig. 9 represents the absorption curve before and after optical transformation.

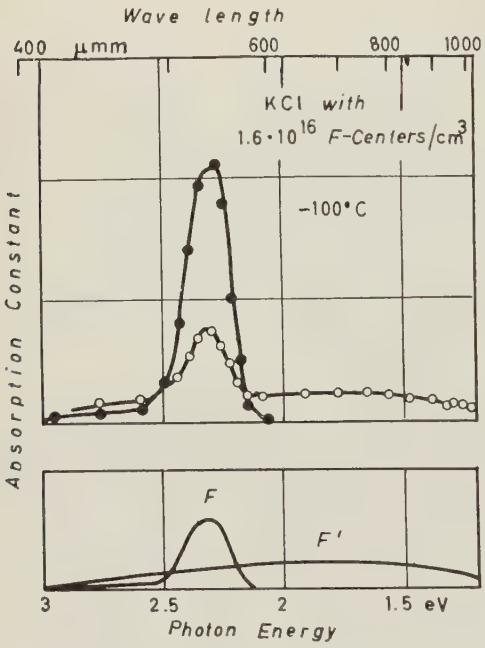


Fig. 9. – The $F \rightarrow F'$ transition in KCl.

The process $F \rightarrow F'$ may be reversed by absorption of light in the spectral range of the F'-band.

The quantum yields as a function of temperature for both processes $F \rightarrow F'$ and $F' \rightarrow F$ are given in Fig. 10. Quantum yield is the ratio: number of electrons transferred/number of light quanta absorbed. The experimental results show maximum values of 2. This gives immediately the following interpretation. The F'-center is formed if an electron is trapped by an F-center. By this process which has been stimulated by *one* optical excitation the state of *two* electrons is changed: the first which has been released and trapped, the second which was present in the trapping F-center and which has to share the vacancy with the newly trapped electron.

Let us consider the two reactions:

1) The $F \rightarrow F'$ conversion is started with the optical excitation:

$$F + h\nu \rightarrow F^* .$$

Below — 200 °C the probability for the electron to escape from the excited center F^*, that is to be ionized is very small. The probability increases with increasing temperature and becomes 1 at about — 100 °C. The curve of the quantum yield is given by the temperature dependence of the reaction:

$$F^* \xrightarrow{kT} F^+ + e\,.$$

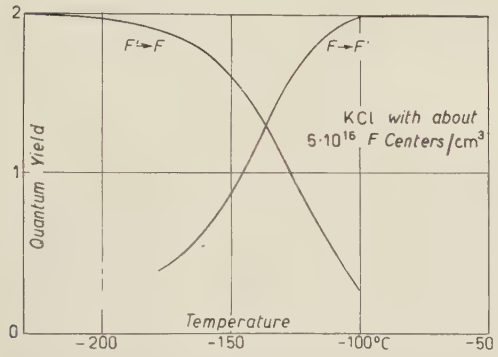


Fig. 10. – The quantum yield of the processes $F \rightarrow F'$ and $F' \rightarrow F$ in KCl.

At the beginning of the process mainly F-centers are available as trapping centers for freed electrons. The process

$$F + e \rightarrow F'$$

has full efficiency in the direction from the left to the right. In the course of time two back-reactions occur. One is caused by the increasing number of ionized F-centers; these are empty vacancies. More and more electrons are retrapped in F-centers:

$$F^+ + e \rightarrow F\,.$$

The other is caused by the overlapping of the absorption bands F and F'. An increasing number of light quanta absorbed in the spectral range of the F-band is caught by the F'-band. This is equivalent to a destruction of F'-centers, corresponding to

$$F' + h\nu \rightarrow F + e\,.$$

The two back reactions account for the saturation value of the $F \rightarrow F'$ conversion (Fig. 11).

2) $F' \rightarrow F$ conversion is started with the optical excitation of the F'-center. This process has full efficiency down to very low temperatures. This probably means that the optical excitation is always equivalent to optical ionization,

$$F' + h\nu \rightarrow F + e\,.$$

The small value of the quantum yield at higher temperatures may be understood by the

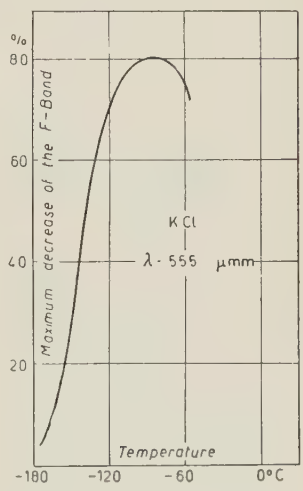


Fig. 11. – The maximum values for the $F \rightarrow F'$ conversion in KCl.

high cross-section for electron trapping of the F-center in comparison with the empty vacancy.

At high temperatures the F'-centers become instable (KCl above — 80 °C). They are decomposed by thermal ionization. The F-centers are still effective as traps, but the lifetime of the F'-binding gets smaller and smaller. So all the electrons are finally in the F-state.

9. – Photoconductivity.

The F-centers as well as the F'-centers are sources for mobile electrons; the F-centers in the temperature range of thermal ionization of optically excited F-centers: the F'-centers probably in the whole temperature range of their existence. In Fig. 12 the photocurrent is plotted against temperature for additively colored KCl [18]. Four sections of the curve may be distinguished:

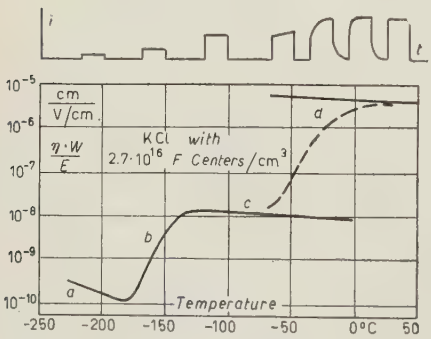


Fig. 12. – Photoconductivity in dependence of temperature: KCl.

a) At very low temperatures (in KCl below — 180 °C) the photoelectric sensitivity is very small. It has not yet been cleared up whether this photocurrent is given by a small amount of optical ionization of F-centers or by other imperfections.

b) Within a narrow range of temperature a steep rise of sensitivity is observed.

c) It is followed by a flat region with a slow falling-off which may be followed up to high temperatures.

All these effects have a very small time constant. The time to establish the equilibrium between production and trapping of mobile carriers is less than 10^{-5} s [19].

d) In the same temperature range a second value of the photoconductivity may be observed. The primary photocurrent which rises instantaneously as the light is switched on is superimposed on a secondary one, which increases slowly with time. At — 50 °C the time constant is of the order of 100 s. It decreases rapidly with increasing temperature. It depends on the time constant of the instrument used with which one of these photocurrents is measured.

At temperatures above 200 °C the photocurrent is covered by the ionic conductivity of the material. It should be mentioned that in the total temperature range below room temperature the crystal in the dark is an excellent

insulator, that means there are no carriers present. If a photocurrent is flowing there is no transfer of carriers from the electrodes to the crystal. As a result the photocurrent is polarized in a very short time, as soon as

$$Q \cdot l/d = U ,$$

where: Q = displaced charge,
l = displacement,
d = thickness of the crystal,
U = voltage.

To measure the photocurrent only very small quantities of charge should be displaced.

The interpretation of the photoconductive behaviour of the crystal with F-centers may easily be given as follows:

The steep rise in the temperature range from -180 °C to -130 °C is given by the increase from zero to one in probability for thermal ionization of excited F-centers. The activation energy needed is of the order of 0.02 eV. Above -130 °C any optically excited F-electron will be freed. Above this temperature the quantum yield η for the production of mobile electrons is one. The soft decrease of the curve with increasing temperature may be explained by the temperature dependence of the mobility v.

The first build up of the equilibrium of the photocurrent is given by the optical-thermal ionization on one side, the trapping of electrons by F-centers on the other side, both these processes happening with a very short time constant. The binding of electrons in F'-centers now may be resolved by thermal activation. The trapped electron may again be freed. These processes of trapping mobile electrons as F'-centers and freeing by thermal activation are repeated again and again until the electron is trapped as a stable F-center. The time constant of the secondary photocurrent is mainly given by the lifetime of the F'-binding. This time is long at low temperatures and small at high temperatures. The electron being freed several times before it gets retrapped as a stable F-center, the photocurrent d is several times higher compared with the photocurrent c in Fig. 14. To give quantitative values for the photoelectric sensitivity the following expression may be used:

$$\eta \overset{w}{}_{E} = v \cdot \tau ,$$

where: η = quantum yield,
w = displacement distance,
E = field strength,
v = mobility,
τ = mean life time of mobility.

TABLE III.

NaCl	KCl	RbCl	KBr	RbBr	KI	
5.0	2.5	1.0	1.5	2.0	0.6	$\cdot 10^{-8} \frac{\text{cm}}{\text{volt/cm}}$

$\eta w/E$ in the temperature range of primary currents. All values have been reduced to an F-center concentration of $10^{16}/\text{cm}^3$.

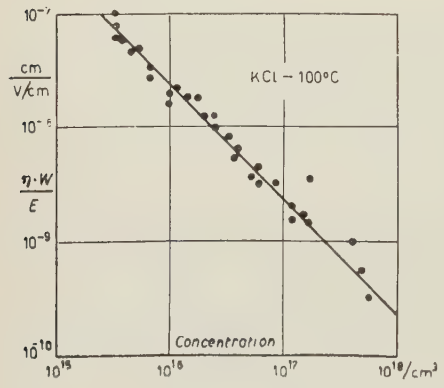


Fig. 13. – Photoconductivity in dependence on F-center concentration in KCl.

Table III gives a few values of $\eta w/E$ in the temperature range of full quantum yield.

The value $\eta w/E$ depends on the concentration of F-centers. As Fig. 13 shows $\eta w/E$ is proportional to the inverse concentration of F-centers. This means the mobile electrons mainly move randomly, they diffuse through the crystal. With an applied electric field the component of the average free path is somewhat longer in the direction of the field. At a field strength of 1000 V/cm and F-center concentration of $10^{16}/\text{cm}^3$ w is of the order of 0.2 μm.

10. – Fluorescence of excited F-centers.

At low temperatures optically excited F-centers are transferred to the ground state by fluorescent emission. The experimental results [20] may be given in the following figures:

— Fig. 14 shows the spectral distribution of the emitted fluorescence light for different temperatures. The maximum of the emission curve is near 1.1 μm in the infrared. This means a quantum loss of about 50% (the maximum of the absorption curve being near 0.55 μm). The fluorescence emission disappears above — 50 °C.

— Fig. 15 summarizes the quantum efficiency of the fluorescence in dependence on temperature for a crystal with small concentration of F-centers. At low temperatures we observe a quantum efficiency of the order of 1. It drops down rapidly at temperatures above — 160 °C.

— Fig. 16 gives the dependence of quantum efficiency on the F-center concentration at low temperature. The quantum efficiency decreases proportionally to $1/r$, with r the mean distance of the F-centers.

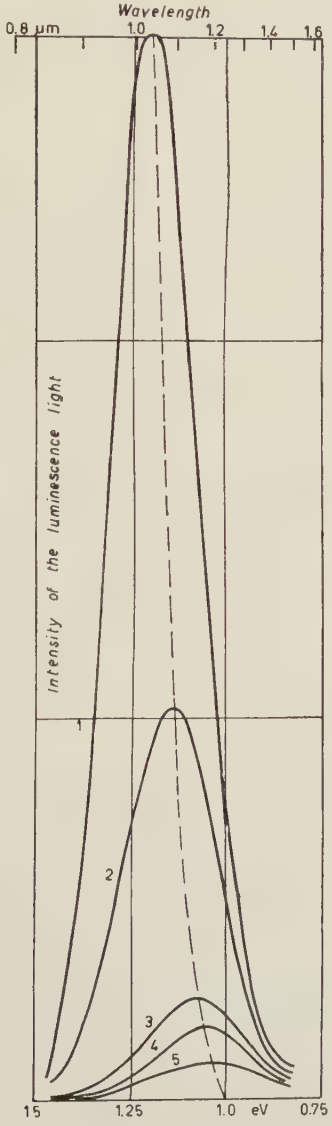


Fig. 14. – Spectral distribution of F-center fluorescence in KCl at different temperatures. The scale is the same for the different curves, but there is no correction concerning dispersion of the prism used.

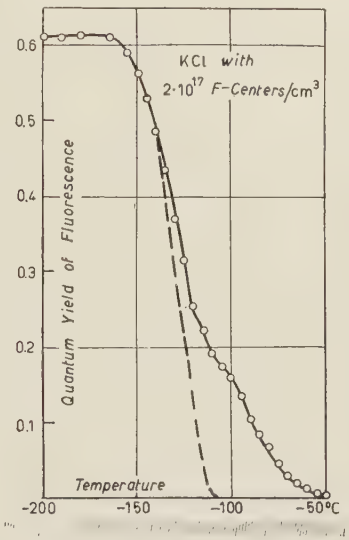


Fig. 15. – Temperature dependence of quantum efficiency of F-center fluorescence in KCl.

The interpretation may be given in the following way: The optical excitation takes place at the microstructure of the center in the ground state. The electronic transition is followed by

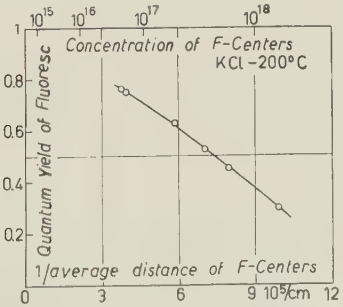


Fig. 16. – Quantum efficiency of fluorescence in dependence on F-center concentration in KCl.

reorientation of the ions. So 50% of the excitation energy is used to establish this lattice polarization. The energy loss gives the Stokes shift.

If there is no thermal ionization during the life-time of the excited state, we find fluorescence emission. With increasing temperature the probability for thermal ionization of the excited center increases rapidly.

A good interpretation for the dependence of the fluorescence efficiency on concentration has not yet been given. So a discussion at this time is inappropriate.

Another type of luminescence has to be mentioned. Many of the F-centers may have been transformed into F'-centers (in KCl at — 100 °C). The crystal is cooled down to — 180 °C. At this temperature the electrons are reversed to form F-centers by optical excitation in the spectral range of the F'-band. The photocurrent remains very small but the process $F' \rightarrow F$ has full quantum yield. The electron being trapped in the empty vacancy to form an F-center in the ground state emits infrared light. The efficiency of this emission decreases rapidly with temperature quite similar to that of the F-fluorescence. The value of the quantum efficiency at low temperatures is of the order of 1. This seems to be a very important result: The F-center retains the ability of light emission if it has been ionized. The spectral distribution of this emission has not yet been measured. The discussion of these emission effects is still going on and is not yet fully understood.

11. – Radiationless transition.

An F-center being excited optically always transfers excitation energy to the lattice as heat. Above — 50 °C there is no fluorescence and no luminescence. In addition there is no storage of energy in metastable states. In this temperature range all the excitation energy is transferred into heat. At low temperatures (as — 180 °C) part of the excitation energy is emitted as fluorescence. Under favorable conditions one observes in KCl a quantum efficiency of 75%. As a result 25% of the excited centers go to the ground state by a radiationless transition. In addition to this energy loss there is a quantum loss. From the Stokes shift one deduces a loss of 50%. This means in KCl at — 180 °C

$$0.25 + 0.75 \cdot 0.5 = 63\,\%$$

of the absorbed energy should be transferred into heat. As a result of calorimetric investigations a value of 61% was found [21].

12. – Field ionization of excited F-centers.

It was discussed in Sect. **7** that an optically excited F-center may be ionized if the temperature is high enough. As a result of the ionization the photoconductivity and the $F \rightarrow F'$ transition were observed. From the temperature of full quantum yield (in KCl at -100 °C) the ionization energy can be estimated. It is of the order of 10^{-2} eV. The energy being rather

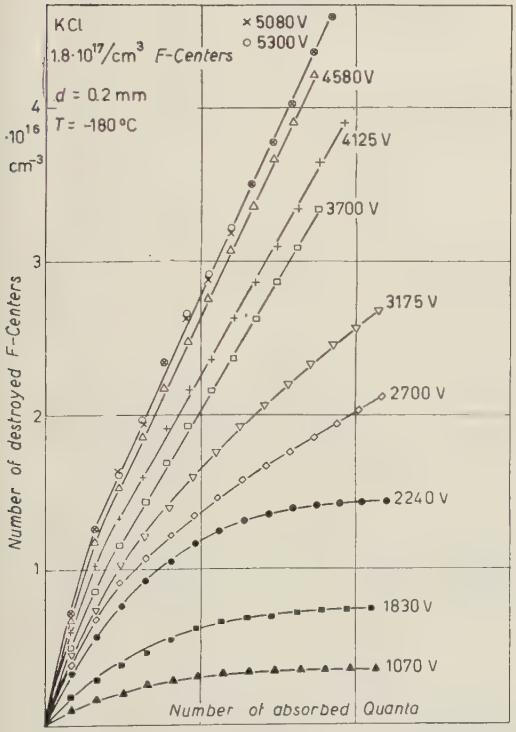


Fig. 17. – The F-center reduction by light absorbed in the F-band as a function of time for different applied voltages (KCl).

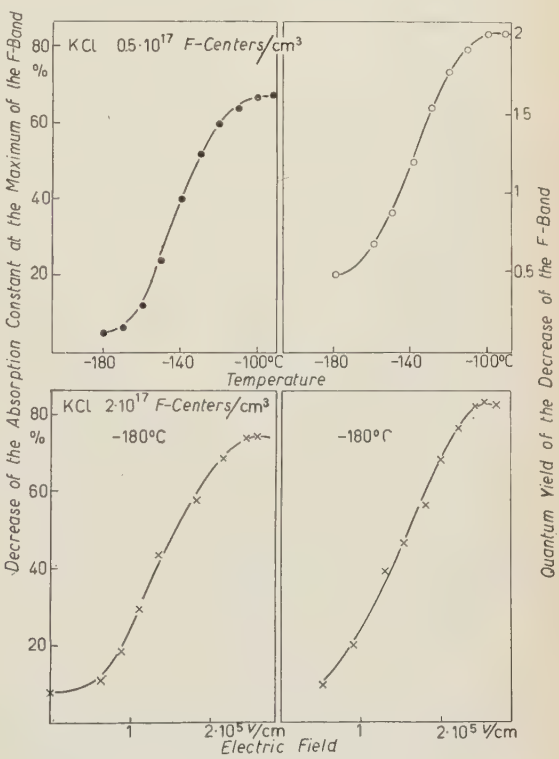


Fig. 18. – Quantum efficiency and F-reduction as a function of applied electric field and of temperature (KCl).

small there is a chance to do the same by the energy of an applied electric field. As a result of such experiments one finds the following [22]:

— In Fig. 17 the reduction of the F-absorption of KCl with the time under excitation with light of the F-band at -180 °C at different applied voltages. One realizes that the $F \rightarrow F'$ transition is favoured by an electric field. The initial slope of these curves gives the quantum yield of the process.

— In Fig. 18 the dependence of the quantum yield on field strength is plotted and may be compared with the dependence on temperature (right hand). The left hand curves give the same dependences for the maximum values of F-reduction or F'-production.

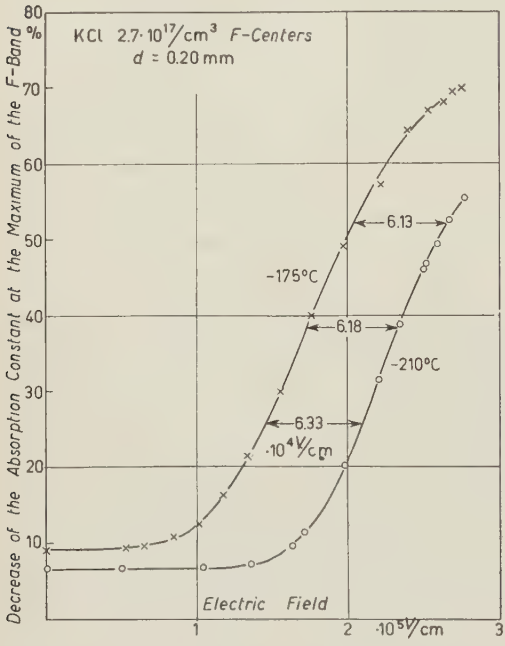


Fig. 19. – F-reduction as a function of the applied field for two different temperatures (KCl).

— In Fig. 19 both the heat and the field energy may be compared. The curves represent the F-reduction as a function of voltage for two different temperatures. The temperature difference of 35° corresponds to an electric field of $6.2 \cdot 10^4$ V/cm.

From

$$kT = \frac{eEa}{\varepsilon},$$

one derives for a, the distance in the direction of the field in which the F-electron is able to receive field energy a value of 15 Å. It is the same order of magnitude as the lattice constant. This order has the meaning of an estimate. The local field is probably different from the applied field.

13. – The paramagnetic moment of the F-center.

It has been known for a long time [23] that an alkali halide crystal becomes paramagnetic by the presence of F-centers. As may be seen in Fig. 20 the increase in magnetic susceptibility is proportional to the concentration of F-centers. The paramagnetism is reduced by the transformation of F-centers into F'-centers. This is in good agreement with the model of the F-center.

A more precise analysis of the electronic structure of the F-center may be obtained by the methods of high frequency spectroscopy, the observation of the electron spin resonance [24]. The result is the following: At a high frequency of the electric field of $9 \cdot 10^9$/s a KCl-crystal with $(10^{17} \div 10^{18})$ F-centers/cm³ shows an absorption line at a magnetic field of 3000 Oe with a line

width of 54 Oe. The absorption band has a Gaussian distribution. The g-value is 1.995.

The g-value is a reference to the unpaired electron. The very broad absorption line cannot be understood either as a result of a dipole-dipole coupling of neighbouring F-centers or as unpaired electrons in a dielectric medium. Then the line width would be of the order of 0.1 or several Oe. A better result may be obtained by the suggestion of the magnetic interaction of electrons and nucleons. In a first order calculation one has to consider the six nearest neighbours surrounding the vacancy. These are the K^+ ions with nuclear spin $\frac{3}{2}$. There are four orientations possible: $\frac{3}{2}$; $\frac{1}{2}$; $-\frac{1}{2}$; $-\frac{3}{2}$. Compared with an external magnetic field, with 6 atoms, this means $4^6 = 4\,096$ different combinations. If the orbit of the F-electron is a linear combination of potassium orbits then the paramagnetic resonance curve is the envelope of all these 4 096 combinations. The electron being of pure s-type one would expect 92 Oe for the line width. The observation of 54 Oe is understood as an admixture of about 40% p-type wave functions of the electron. The reason for the difference from the s-type orbit is probably given by the polarizing influence of the vacancy.

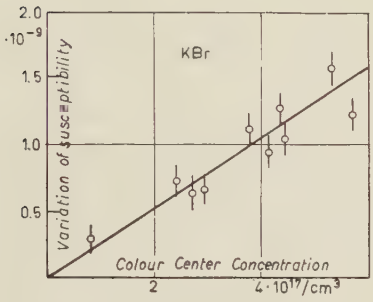


Fig. 20. – The paramagnetic susceptibility of KBr-crystals with different concentrations of F-centers.

There is an interesting result with KCl which has a nearly pure isotope ^{41}K. The magnetic moment of ^{41}K is only 0.563 of that of ^{40}K. The half width should therefore be reduced to 31 Oe. The experiment gives 36 Oe. As a result one suggests the co-operation of the next nearest neighbours, the 12 Cl^--ions. 1/6 of the electronic density has to be attached to the chlorine ions.

All these results are in very good agreement with the model of the F-center, but they give a more precise idea of the electronic ground state. The remark should be added that recently the separation of the single lines within the broad resonance band could be found by a double resonance method [25].

14. – The diagram of the configurational coordinates.

To summarize the most important results it may be useful to consider the diagram of the configurational co-ordinates. In the schematic diagram of Fig. 21 the meaning of the abscissa is the distance between ions surrounding the F-center electron. The curves are drawn for the negative ions. The ordinate is the potential energy.

Curve F represents the energy of the ground state of the F-center. The minimum corresponds to the equilibrium positions in a static lattice, with no vibrations. The zero point energy is given by a level $\frac{1}{2}\hbar\omega_{\parallel}$ above the minimum. The different vibrational states at higher temperature then might be represented by levels with constant distance $\hbar\omega_{\parallel}$. So the range of configurational co-ordinate realized at a given temperature is given by the number of excited vibrational quanta.

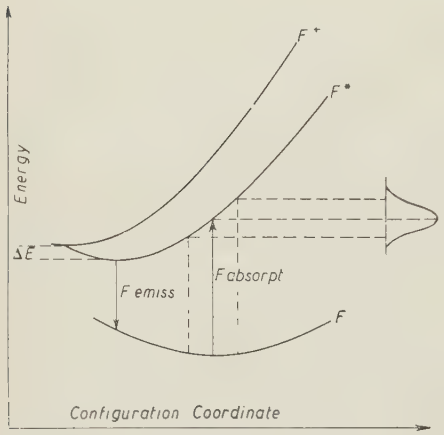


Fig. 21. – Energy level diagram.

Curve F^* represents the energy of the excited state. The electron of the F-center spreads out over a wider range of the crystal; the negative ions surrounding the vacancy have to move nearer to the center of the vacancy in order to produce an energetic minimum. So the curve F^* is shifted towards smaller values of the configurational co-ordinates.

For the same reason the minimum of the ionized state F^+ again is given at smaller distances of the negative ions. (If the co-ordinates of the positive ions are plotted, the energy curves are shifted in the opposite direction.)

The optical absorption transfers the F-center from the ground state to the excited state. This state is unstable and goes to its energetic minimum by reorientation of the ions. The excess in energy is used for the emission of phonons, that means to heat the crystal. A similar process happens if the center goes from the minimum of the excited state to the ground state after emission of fluorescence light. The sum of these two losses of energy of optical transition is observed as Stokes-shift of the fluorescence. The temperature dependence of the maximum of the F-band and the half width of the band may easily be understood with the help of this diagram.

Special interest might be given to the intersection of the potential curves. Such an intersection is realized at high enough vibrational states, that means at high enough temperatures. At these temperatures radiationless transitions from excited states to the ground state may be realized. But there is no chance to make predictions without precise knowledge of the potential curves. To get this information the theory of PEKAR has to be developed to the treatment of the movement of the single ions of the imperfections, but this has not yet been done.

REFERENCES

[1] E. Goldstein: *Zeits. f. Instrumentenkunde*, **16**, 211 (1896).

[2] Z. Gyulai: *Zeits. f. Phys.*, **37**, 889 (1926); see: H. Rose: *Pogg. Ann.*, **120**, 1 (1863); F. Giesel: *Ber. d. Deutsch. Chem. Ges.*, **30**, 156 (1897); H. Siedentopf: *Phys. Zeits.*, **6**, 855 (1905); Z. Gyulai: *Zeits. f. Phys.*, **35**, 411 (1926).

[3] I. Alexander and E. E. Schneider: *Nature, London*, **164**, 655 (1949); H. Dorendorf and H. Pick: *Zeits. f. Phys.*, **128**, 166 (1950); H. Dorendorf: *Zeits. f. Phys.*, **129**, 317 (1951); R. Casler, P. Pringsheim and Ph. Yuster: *Journ. Chem. Phys.*, **18**, 887, 1564 (1950).

[4] E. Mollwo: *Nachr. Akad. Wiss. Göttingen*, **1935**, 215; *Ann. der Phys.*, **29**, 394 (1937).

[5] G. Tammann and R. Veszi: *Zeits. anorg. u. allg. Chem.*, **150**, 355 (1925); S. Kyropoulos: *Zeits. anorg. u. all. Chem.*, **154**, 308 (1926).

[6] R. W. Pohl: *Proc. Phys. Soc.*, **49**, 3 (1937); *Phys. Zeits.*, **39**, 36 (1938).

[7] F. Seitz: *Rev. Mod. Phys.*, **18**, 384 (1946); **26**, 7 (1954).

[8] W. Martienssen and H. Pick: *Nachr. Akad. Wiss. Göttingen*, **1952**, 17; *Zeits. f. Phys.*, **135**, 309 (1953); W. Martienssen and H. Mette: *Naturw.*, **41**, 331 (1954); G. Miessner: *Zeits. f. Phys.*, **141**, 111 (1955).

[9] H. Rögener: *Ann. der Phys.*, **29**, 386 (1937).

[10] O. Stasiv: *Nachr. Ges. Wiss. Göttingen*, **1932**, 261; G. Heiland: *Zeits. f. Phys.*, **127**, 144 (1950).

[11] E. Mollwo: *Nachr. Ges. Wiss. Göttingen*, **1939**, 149.

[12] J. H. de Boer: *Recueil des Trav. Chim. d. Pays-Bas*, **56**, 301 (1937).

[13] S. I. Pekar: *Journ. Phys. USSR*, **10**, 341, 347 (1946); *Žu. Éksper. Teor. Fiz.*, **19**, 796 (1949). *Untersuchungen über die Elektronentheorie der Kristalle.* Akademie Verlag (Berlin, 1954) (Translation of a russian publication in 1951); *Usp. Fiz. Nauk*, **50**, 197 (1953); see: K. Huang and A. Rhys: *Proc. Roy. Soc.*, A **204**, 406 (1950).

[14*a*] F. Stöckmann: *Naturw.*, **39**, 230 (1952).

[14*b*] E. Mollwo: *Nachr. Ges. Wiss. Göttingen*, **1931**, 97; H. F. Ivey: *Phys. Rev.*, **72**, 341 (1947).

[15] E. Mollwo: *Zeits. f. Phys.*, **85**, 56 (1933).

[16] A. Smakula: *Zeits. f. Phys.*, **59**, 603 (1930).

[16*a*] F. G. Kleinschrod: *Ann. der Phys.*, (5) **27**, 97 (1936); R. H. Silsbee: *Phys. Rev.*, **103**, 1675 (1956); C. J. Rauch and C. V. Heer: *Phys. Rev.*, **105**, 914 (1957).

[17] R. Hilsch and R. W. Pohl: *Zeits. f. Phys.* **68**, 721 (1931); H. Pick: *Ann. der Phys.* **31**, 365 (1938).

[18] G. Glaser and W. Lehfeldt: *Nachr. Ges. Wiss. Göttingen*, **2**, 91 (1936); G. Glaser: *Nachr. Ges. Wiss. Göttingen*, **2**, 109 (1936); **3**, 31 (1937); *Ann. der Phys.* (5), **27**, 217 (1937); A. Redfield: *Phys. Rev.*, **91**, 244, 755 (1953).

[19] O. v. Guericke: *Nachr. Wiss. Akad. Göttingen*, **1950**, 1.

[20] Th. P. I. Botden, C. Z. van Doorn and Y. Haven: *Philips Res. Rep.*, **9**, 469 (1954); K. H. Becker and H. Pick: *Nachr. Akad. Wiss. Göttingen*, **1956**, 167 and unpublished results.

[21] E. Feldtkeller: *J. Phys. Chem. Solids* (1957), to be published.

[22] F. Lüty: *Color Center Symposium at Argonne Nat. Lab.*, November 1 (1956).
[23] P. Jensen: *Ann. der Phys.*, (5) **34**, 161 (1939).
[24] C. A. Hutchison: *Phys. Rev.*, **75**, 1769 (1949); E. E. Schneider and T. S. England: *Physica*, **17**, 221 (1951); C. A. Hutchison and G. A. Noble: *Phys. Rev.*, **87**, 1125 (1952); A. H. Kahn and C. Kittel: *Phys. Rev.*, **89**, 315 (1953); C. Kittel and E. Abrahams: *Phys. Rev.*, **90**, 238 (1953); A. F. Kip, C. Kittel, H. A. Levy and A. M. Portis: *Phys. Rev.*, **91**, 1066 (1953); A. M. Portis: *Phys. Rev.*, **91**, 1071 (1953); E. E. Schneider: *Phys. Rev.*, **93**, 919 (1954).
[25] G. Feher: *Phys. Rev.*, **105**, 1122 (1957); N. W. Lord: *Phys. Rev.*, **105**, 757 (1957).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Radiation Effects in Solids (*) (+).

R. SMOLUCHOWSKI

Carnegie Institute of Technology - Pittsburgh, Pennsylvania

1. – Introduction.

A systematic study of irradiation effects in solids started during the war although such effects were known before (photographic effect, influence of α-particles on certain minerals and associated coloration phenomena).

Historically the first suggestion concerning nuclear irradiation effects was made by WIGNER during the war, in connection with the construction of reactors. He pointed out that properties of graphite, which are important from the engineering point of view, will be strongly affected by nuclear radiation.

At present the motivation for the study of irradiation effects is both basic and applied. The applied problems are obvious in power reactors where there are fuel elements, moderator, coolant, etc., with materials such as uranium, beryllium, beryllium oxide, uranium oxide, mica, steel, aluminum. From the basic point of view the interest lies in the quantitative understanding of the elementary phenomena and in the analysis of the secondary effects.

The progress in the fundamental studies and in practical applications is still fairly independent. There is a tremendous amount of work being done on engineering materials which is usually very difficult to understand from a scientific point of view.

The theory of irradiation effects has two main aspects:

a) Collision theory as applied to solids.

b) Theory of the influence of defects upon various properties of solids.

The first part, the collision theory, is probably in a better shape than the second.

(*) Supported by a A.E.C. research contract.

(+) This text is based on Lecture Notes collected by a group of students (A. ASCOLI, M. FUJIMOTO and F. A. LEVI) and revised by the author.

The plan of these lectures is to give first a brief summary of the theory and then to illustrate the various phenomena with a few examples chosen from the domain of metals and non-metals.

2. – Basic phenomena.

Particles used in irradiation studies are α-particles, protons, deuterons, neutrons, electrons, fission fragments and γ-rays. From the engineering point of view neutrons and fission fragments are the most important, but generally we do not know well enough their velocity spectrum or their flux, so that a control and interpretation of the experiments in piles is difficult. The situation is much better with particles accelerated in cyclotrons, Van de Graaffs etc.

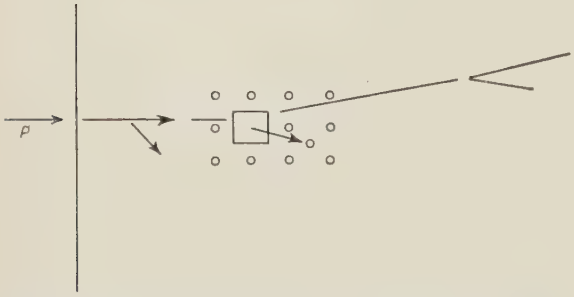


Fig. 1. – Defect production by an incident particle.

What happens when a charged particle, for instance a proton, enters a solid, may be pictured, Fig. 1, in the following way: first of all, being charged, it produces a large amount of electronic excitation and ionization, then it may also make elastic collisions with the atoms of the lattice giving to them part of its energy. If the imparted energy is rather small it results in an increased vibration of the atoms of the lattice; if it is high it may displace an atom from its original position so that it moves through the lattice and eventually becomes an interstitial. This primary displaced atom may on its way collide with other atoms producing secondaries, tertiaries etc. The energy losses of a typical incident fast proton are roughly distributed in the following way:

Electron excitation	99.9 %
Elastic collision	.05 %
Displacements	.05 %

At high energies also inelastic collisions may happen: the incident particle is absorbed by a nucleus which subsequently disintegrates. Transmutation by thermal neutrons should be mentioned too. Summarising we have:

1) Electronic excitation.
2) Elastic collisions.
3) Displacements.
4) Inelastic collisions
5) Transmutations (thermal neutrons).

3. – Elastic collision.

A charged particle approaching an atom may produce a Coulomb encounter, the maximum energy transmitted being given by

$$T_m = \frac{4M_1M_2}{(M_1+M_2)^2}E\,,$$

where M are masses and E the energy of the incident particle; here as in the following lectures, the subscript 1 is used for the incident particle, and the subscript 2 for the stationary atom.

The minimum energy required to displace an atom from its position in the lattice may be taken to be about $E_d = 25$ eV (SEITZ). Typical energies of incident particles necessary to displace an atom of given mass are given in the following table:

Incident particle	$M_2 = 10$	$M_2 = 100$
Electron	$1 \cdot 10^5$ eV	$68 \cdot 10^5$ eV
Proton	76	638
α-Particle	31	169
Fission fragment . . .	76	25

Here the values for electrons were computed using a relativistic formula, because relativistic energies are involved.

4. – Rutherford scattering.

When a particle of charge e_1 passes with velocity V in proximity of another particle of charge e_2, it deviates by an angle θ, p being the impact parameter, Fig. 2.

The distance of closest approach, b, is given by the relationship

$$\frac{e_1e_2}{b} = \frac{\mu V^2}{2}\,,$$

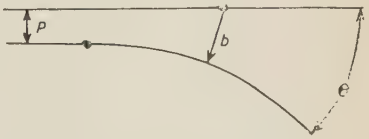


Fig. 2. – Rutherford scattering.

μ being the reduced mass. The differential cross-section for Rutherford scattering is given by

$$d\sigma = \frac{\pi b^2}{4} T_m \frac{dT}{T^2} ,$$

where T_m is the maximum transmitted energy, T the kinetic energy actually transmitted. It is important to realize that here $d\sigma$ is dependent on $1/T^2$, which means that we have a peak at low energies. This is not the case for neutrons as we shall see later. The Rutherford formula is valid also in the quantum mechanical range, if b is bigger than the wavelength

$$b \gg \lambda = \frac{h}{\mu V} .$$

5. – Screening.

Since we are dealing with atoms, we must take into account the screening of the nuclear charge by the electrons. The corresponding potential is given by:

$$\frac{e_1 e_2}{r} \exp\left[-\frac{r}{a}\right] .$$

Using the Thomas-Fermi model of the atom, the screening constant a is

$$a = c \frac{a_h}{Z_2^{\frac{1}{3}}} ,$$

where a_h is the Bohr radius, Z_2 is the atomic number of the stationary atom and c is of the order of unity.

If the screening of the charge of the incident particle is to be taken into account we may use the expression suggested by BOHR:

$$a = \frac{a_h}{\sqrt{Z_1^{\frac{2}{3}} + Z_2^{\frac{2}{3}}}} ,$$

which, for $Z_1 = Z_2$, reduces to

$$\frac{a_h}{\sqrt{2} Z_2^{\frac{1}{3}}} .$$

BRINKMAN in his calculations used instead the expression

$$\frac{a_h}{Z_2^{\frac{1}{3}}} .$$

The factor $\sqrt{2}$ changes considerably the amount of effective screening present when two atoms collide.

Let us distinguish between weak and strong screening, weak screening being when $b \ll a$, and strong screening when $b \gg a$. If we have a *weak screening*, and if $\lambda \ll b \ll a$, classical Rutherford scattering is applicable down to small angles of the order of $\theta'_a \cong b/a$; if, however, $\lambda \gg b \ll a$, quantum mechanics has to be used, Rutherford scattering being valid only for angles bigger than $\theta''_a \sim \lambda/a$. These are the two limiting values of θ for weak screening, that is to say when the electrons do not surround the nucleus very tightly. If we have *strong screening*, that is when $b \gg a$, then if also $\lambda \ll a$, the classical hard sphere approximation may be used. Fortunately this is a very frequent case of general applicability. For $\lambda > a$ two possibilities occur: For $\lambda \gg \sqrt{ab}$ the Bohr approximation is applicable, for $\lambda < \sqrt{ab}$ the case is very difficult to treat, but fortunately it does not occur in practice in the application to irradiation effects.

The energy transmitted at the two limiting angles θ'_a and θ''_a is given by

$$T'_a \cong 4 Z_1^2 Z_2^2 \frac{R}{E} \frac{M_1}{M_2} \left(\frac{a_1}{a}\right)^2 R_1 ,$$

and

$$T''_a \cong \frac{m_e}{M_2} \left(\frac{a}{a}\right)^2 R_h ,$$

respectively, where R_h is the Rydberg energy and m_e the electron mass. These energies come out to be of the order of 0.01 eV, and therefore are much too small to cause displacements. It follows that Rutherford scattering formulae are applicable in all cases as far as displacements are concerned.

6. – Knock-ons.

Let us see now what happens to the primary displaced atom. The maximum energy transmitted may be expected to be below 10^4 eV. Usually $b \gg \lambda$ for light and heavy atoms. But $b > a$ for intermediate and higher values of Z_2, as for instance for $Z_2 = 30$; $b < a$ for small Z_2. This means that we have a weak screening for high energies E and low atomic numbers Z which is rather an exceptional case. In typical cases we have strong screening. For the majority of cases the hard sphere approximation is satisfactory; only in the initial part of the path of the particle, Rutherford calculation may be necessary, but as soon as the particle slows down the hard sphere approximation is sufficient.

A light displaced atom in a solid starts with Rutherford collisions ending with hard sphere collisions which may be looked upon as a slow diffusion process. Heavy atoms, on the other hand, diffuse from the beginning.

7. – Cross section for incident particle.

The total cross-section for Coulomb encounters is given by

$$\sigma_a = \begin{cases} \pi a^2 & \text{for } b > \lambda \\ \dfrac{\pi a^2 b^2}{\lambda^2} & \text{for } b < \lambda \,, \end{cases}$$

while the total cross section for displacements is given by

$$\sigma_d = 4\pi a_h^2 \frac{M_1}{M_2} \frac{Z_1^2 Z_2^2 R_h^2}{E E_d} \,,$$

which is a fundamental formula for all calculations on radiation problems.

From this we may calculate the mean energy transmitted,

$$\overline{T}_d = E_d \log \frac{T_m}{E_d} \qquad (T_m \gg E_d) \,,$$

while for Coulomb encounters

$$\overline{T}_a = T_a \log \frac{T_m}{T_a} \,.$$

8. – Mean free path. Displacement spikes.

Knowing the cross-section for Coulomb collisions, we can now calculate the mean free path of a particle

$$L_a = \frac{1}{n_0 \sigma_a} \,,$$

where n_0 is the number of atoms per cm^3.

For $b > \lambda$, $\sigma_a \cong 10^{-16}$ cm^2 for small Z_2, σ_a becomes even smaller for increasing atomic number. For $b < \lambda$, $\sigma_a < \pi a^2$ and it increases with Z_2.

According to the values assumed by σ we may then have either local regions of damage distributed along the path of the particle or the regions may merge into an elongated damaged zone.

We may speak of « displacement spikes » when a large number of atoms are displaced in the same region. An important difference between thermal spikes and displacement spikes is that the latter are big ($10^4 \div 10^5$ atoms are « heated » to above the melting point) and involve a lot of energy, ($E \cong \overline{T}_d$),

giving rise to vacancies and interstitials, their life being of the order of 10^{-11} s. Thermal spikes, discussed in more detail later on, are smaller (less than 100 atoms), the energy E being of the order of 10 eV, with no interstitials and vacancies produced and a life time of the order of 10^{-12} s.

If the hard sphere model is used, the important question arises what is the effective radius R of the sphere. If the screening potential is proportional to $\exp[-r/a]$, then

$$R \cong a_h \qquad \text{for} \quad Z_2 > 10 \qquad \text{and} \quad E \gg E_d \,.$$

In the case of small energy ($E < E_d$), R may increase up to the actual radius of the atom in the lattice r_s.

Calculating the mean free path in the hard sphere approximation we get, for copper,

$$L_{\text{hard sphere}} \cong 6\, r_s \,.$$

The energy of the displaced atom is then of the order of 500 eV.

It should be borne in mind that several different approximations may be used; in particular BRINKMAN calculated the total energy transfer if two atoms encounter at a given distance. He obtained appreciably smaller values for L and in particular cases the mean free path appears to be as small as $0.5\, r_s$, that is smaller than atomic dimensions. This consideration is very important with reference to the structure of displacement spikes.

9. – Energy loss by incident particle from Coulomb collisions.

The rate of energy loss by Coulomb collisions is given by

$$\left.\frac{\partial E}{\partial x}\right|_{\text{coll.}} = 2\pi n_0 \frac{Z_1^2 Z_2^2 e^4}{M_1 V^2} \log \frac{T_m}{T_a} \,,$$

which indicates that the rate of loss of energy is inversely proportional to the energy.

10. – Electron irradiation.

In the case of electron irradiation, the proper theory is relativistic, and quantum mechanical calculations have to be made. The screening is always weak. Usually no secondaries are produced and thus a comparison of experiment with theory is rather simple.

11. – Neutron irradiation.

Irradiating with neutrons, the maximum energy transferred in a collision with an atom of atomic weight A_2 is given by

$$T_m = \frac{4A_2}{(1+A_2)^2} E ,$$

the differential cross-section (isotropic) is given by

$$\mathrm{d}\sigma = \sigma_s \frac{\mathrm{d}T}{T_m} ,$$

σ_s being the total cross-section. In contrast to Rutherford scattering there is no peak for low energies, most collisions transmit a high amount of energy, and therefore secondaries and tertiaries are very important in estimating the total number of displacements. The mean free path for neutrons in a typical case is of the order of 1 cm, which is so long that independent regions of energy release are created.

12. – Electronic excitation.

Moving atoms in a solid lose those electrons which have an orbital velocity lower than the velocity of the atom itself. In this way the effective charge of the ion changes with velocity: the faster the atom moves, the more it is stripped of its electrons.

If we consider electronic excitation in metals, we have to take into account the Fermi distribution and then generally only the outside electrons may be excited. In the calculations it is convenient to use the reduced energy $\varepsilon = (m/M)E$, which is the energy of an electron (mass m) having the same velocity as the atom (mass M). The values of ε of primaries produced by neutrons are usually quite small compared with the Fermi energy.

Total amount of losses per unit length of path due to electronic excitation is given by

$$-\left.\frac{\partial \varepsilon}{\partial x}\right|_{\mathrm{el}} = 4\pi n_0 \frac{Z_1^2 e^4}{m V^2} Z_2' \log \frac{m_2 V^2}{\bar{E}} ,$$

where n_0 is the density of atoms, and Z_2' is the number of electrons in the stationary atoms which can be excited. $\bar{E}$ is the mean effective ionization potential, which may be obtained theoretically or experimentally by transmission measurements. Here again the dependence on energy is of the type 1/energy as it was for elastic collisions.

13. – Thermal spikes and plasticity spikes.

This problem was first treated by BROOKS and by SEITZ.

The temperature as a function of the time t and of the distance r from a spot where energy Q has been released is given by the equation

$$T(r, t) = \frac{Q}{8\pi^{\frac{3}{2}} c\varrho} \cdot \frac{1}{(Dt)^{\frac{3}{2}}} \exp\left[-\frac{r^2}{4Dt}\right],$$

where c is the specific heat, ϱ is the density and D the thermal diffusivity. Typical results of this calculation are as follows: if Q in a big spike is about 300 eV at one point and the melting-point of the material is 1000 °K, then within the time $t_m = 100 t_0 = 10^{-11}$ s ($t_0 \cong 10^{-13}$ s is the time necessary for thermal energy to be transferred from an atom to its neighbour) the temperature will drop down to the melting-point at the radius $r_m \cong 10 r_s$ ($r_s =$ = radius of atom). The material in the spike can be considered either as being melted or as a superheated solid.

A different picture of these phenomena was suggested by BRINKMAN: as mentioned before, according to his conclusions, the mean free path of a displaced atom is of the order of $0.5\, r_s$, so that the primary collision produces a large number of displacements in its immediate vicinity. If the energy released is of the order of (350 ÷ 400) eV, a sizable void is produced in the material. The displaced atoms are pushed, as interstitials, into the surrounding region; and as a result of the pressure the void may subsequently collapse. So far no very convincing evidence for the existence of such catastrophic explosions has been obtained.

SEITZ has made a very careful study of how hot spikes can affect diffusion processes, a problem which is extremely important in the study of solids and in various engineering problems.

Let us assume that a rate process is controlled by a frequency

$$\nu = \nu_0 \exp\left[-\frac{E'}{kT}\right],$$

where ν_0 is the jump frequency, and the activation energy E' is of the order of 1 eV. The number of jumps that can occur during the lifetime of a spike can be calculated from

$$n_j = 0.093 \frac{\nu_0 r_s^2}{D} \left(\frac{Q}{E'}\right)^{\frac{2}{3}}.$$

For a typical case (energy $\sim$ 300 eV, $E' = 1$ eV) n_j is of the order of 4 to 5. These considerations can be applied to order-disorder reactions. The

first studies of such phenomena were carried out on $AuCu_3$. If one assumes that 300 eV is released in a spike and 3 eV is the energy for interchange of two neighbouring atoms, one gets the number of disordering reactions per spike between 10 and 42. To explain the experimentally observed change of resistivity, the number of jumps should be 120. The discrepancy could arise either from a wrong assumption about exchange energy E' or from the fact that migration of interstitials or vacancies should be additionally considered; these, however, would have to migrate for quite a long distance and produce an interchange with about 20 atoms.

The possibility of « replacement collisions » has been suggested by KINCHIN and PEASE; these may take place when the energy of the colliding atom is lower than the displacement energy E_d but sufficient to have one atom pushed out of its place, the moving one being trapped into the vacancy (see Fig. 3). In order to account for the discrepancy, the replacement should travel for about 10 atoms, and this may be too much to expect.

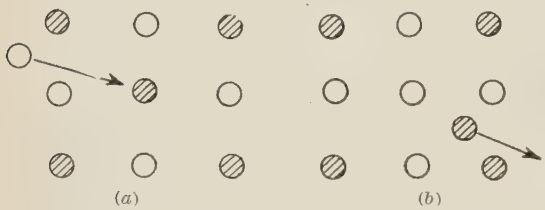


Fig. 3. – Replacement collision.

Another type of spikes, the « plasticity spikes », was introduced by SEITZ, in order to account for the same discrepancy. He points out that the radial stresses surrounding the melted volume may:

a) create new dislocations;

b) move existing dislocations.

The latter possibility *b*) turns out to be unlikely in a well annealed specimen, because the average distance between dislocations and a thermal spike is too big. On the other hand, new dislocation rings could be formed; these would then tend to collapse, the minimum radius for stability being of the order of $5.5\ r_s$. The total energy spent in this process is fairly high, about 30 eV per dislocation loop.

If two such rings cross each other while collapsing, a permanent displacement may be produced, which cannot be removed by slip alone and which contributes to the resistivity. This picture is unfortunately difficult to deal with quantitatively.

Calculating the temperature in a thermal spike, one has to remember that free electrons are excited to a very high temperature and that there is not enough time to establish equilibrium with the lattice. Temperature ratios

are of the order of 200, say 10^5 °K for the electrons and 500 °K for the lattice. If *d*-electrons are excited, one may expect a much shorter relaxation time since the repulsive forces between neighbouring atoms may be directly affected.

14. – The simple theory of displacement.

SEITZ has calculated the total number of atoms displaced per primary, using the model of SNYDER and NEUFELD. This number turns out to be a logarithmic function of the energy, and for all practical purposes it may be assumed that about 5 to 6 atoms are displaced per primary collision. This calculation is based on the hard sphere approximation which, as pointed out before, is usually very well justified.

15. – Inelastic collisions.

Inelastic collisions are important if the energy of the incident particle is high (in the range of 100 to 400 MeV). In this case the incident proton can enter a nucleus, and make one or more direct collisions with some of the nucleons in the nucleus. These will be shot out with energy comparable to the energy of the incident particle. These are the knock-on nucleons.

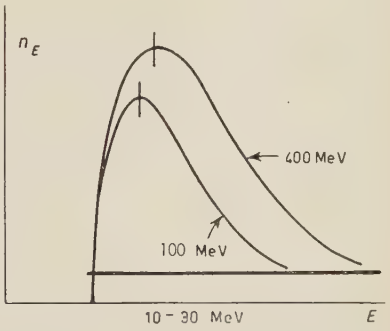


Fig. 4. – Energy spectra of slow nucleons produced by incident protons of 100 and 400 MeV energy.

After one or more such direct collisions within the nucleus have occurred, the nucleus is left in an excited state, and it can be described as a hot liquid drop which cools by evaporation. The excess energy is distributed among the nucleons, so that we have a number of protons, neutrons and deuterons escaping with fairly low energy, of the order of 10 to 30 MeV. The result is that, the primary particle of high energy, by itself not very efficient for producing displacements and electronic excitation, is replaced by a number of slow particles which are very effective in producing defects. Fig. 4 shows the spectrum of energy of the slow emitted particles and the very low intensity of direct knock-ons. As is shown by the two curves the number of

secondary « evaporated » particles increases with the energy of the incident proton, so that, in this case, the total number of defects should go up with the energy of the proton rather than down as in the formulae for elastic and electronic losses.

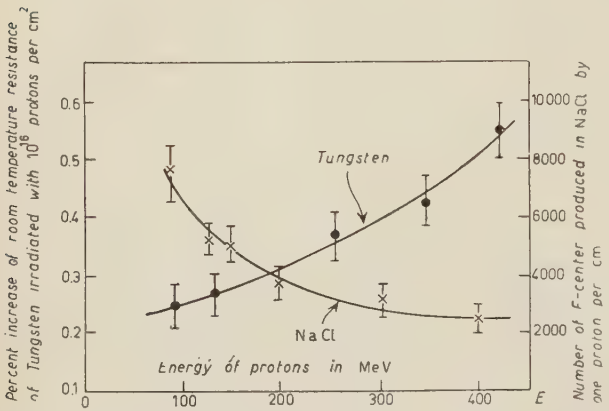


Fig. 5. – Energy dependence of defect formation in W and in NaCl.

This has been checked experimentally by PEARLSTEIN *et al.* by putting a tungsten wire inside a cyclotron, in different positions along its radius, thus changing continuously the energy of the incident protons. Fig. 5 shows the results. The increase of room temperature resistivity of tungsten irradiated with 10^{16} protons per cm^2 has been measured as a function of the energy of the protons and shows a definite slope upwards which is in a quantitative agreement with theoretical estimates.

It contrast with this result one should mention the work of KOBAYASHI on irradiation with high energy protons of a crystal of NaCl in which the primary phenomenon is the formation of Frenkel defects, F centres and others by electronic excitation. Since the energy losses by electronic excitation in this case are about 1000 times higher than all other losses, one expects a $1/E$ dependence on energy. The number of F-centres produced per unit path of the incident proton was measured and is indeed almost exactly proportional to $1/E$. A comparison of the two processes in the same energy range is shown in Fig. 5.

16. – Annealing of defects in metals. The crowdion.

SEITZ has already given in his lectures a very complete account of the theory and experiments pertaining to the production of point defects in metals by irradiation and to their annealing. For this reason here only the related question of a « crowdion » will be discussed.

The annealing of resistivity induced in copper irradiated at 4 °K may be shown plotting, Fig. 6, the resistivity change $\Delta\varrho$ as a function of temperature. It is easily seen that there is a considerable anneal in the range of 30 °K.

SEITZ pointed out that this drop in resistivity is probably due to interstitial migration with a 0.1 eV activation energy.

A possible complication of this picture has been pointed out by BLEWITT *et al.* Adding small amounts of Be or Au to copper the annealing curve is considerably changed: see Fig. 6. There is an almost complete suppression of the annealing phenomena at 30 °K. This seems to speak against the interpretation of the big step as a migration of interstitials since it is difficult to see how a few too small or too big atoms could block the mobility of interstitials. Furthermore the annealing shows a whole spectrum of energies, which is also not easily explained on the basis of interstitial migration. The Oak Ridge group proposed the use of the notion of «crowdion», originally introduced by LOMER and COTTRELL. A crowdion is formed when an interstitial atom relaxes into a line of atoms in a close packed direction, pushing these slightly aside. This arrangement produces a linear defect which may move only along its line (Fig. 7). Inasmuch as impurity atoms and dislocations produce strains which may influence the mobility of a crowdion from a distance, a range of activation energies may be expected for various distances of the impurity atom from the line. A crowdion could also be easily blocked by impurities. It should be pointed out that the very existence of a crowdion as a configuration of lower energy than an interstitial is theoretically as yet unanswered.

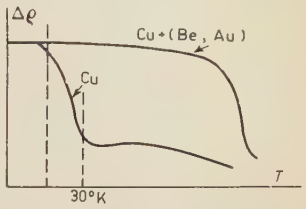


Fig. 6. - Influence of Be and Au additions on the thermal annealing of radiation induced resistivity change in Cu.

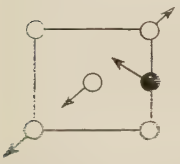


Fig. 7. - Crowdion formation.

17. – Calculations of E_d.

HUNTINGTON, using a simple model with free s and p electrons and a closed d shell, and with a Born-Mayer potential, calculated the values of E_d for various directions of displacement in the lattice of Cu. The atom may be removed either along the cubic diagonal, or along a cube-edge (Fig. 8). In the first case he got $E_d = (18 \div 43)$ eV, in the second $E_d = (14 \div 34)$ eV. His results are not in disagreement with experiments.

Calculations for Ge have been made by KOHN. We shall refer to them later.

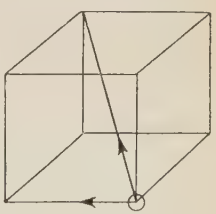


Fig. 8. – Displacement directions in a cubic lattice.

18. – Number of defects.

On the number of defects introduced by irradiation in metals there is a considerable discrepancy between theoretical and experimental values. MARX, HENDERSON and COOPER calculated the change in resistivity for copper irradiated with 10^{17} deuterons at about 100 °K, using Seitz's value of about 6 atoms displaced per one primary, and the early theoretical value of resistivity per defect pair. They found $\Delta\varrho = 1.16\ \mu\Omega$ cm, while the observed value comes out to be $0.23\ \mu\Omega$ cm.

An electron bombardment of Cu was made by EGGEN and LAUBENSTEIN with 810 keV electrons, which is just enough to produce displaced atoms without any secondaries. The theoretical result is in this case $\Delta\varrho = 2.5 \cdot 10^{-8}\ \Omega$ cm, the observed result $\Delta\varrho = 4.4 \cdot 10^{-9}\ \Omega$ cm; so there is again a factor of 5 to 6 between them.

Both theoretical and experimental deductions are based on theoretical values for number of displaced atoms per primary, and on theoretical values for defect contributions to resistivity.

19. – Other effects of irradiation on metals and alloys.

There are various other ways to detect experimentally radiation induced defects in solids. A change in *mechanical properties* was studied at Oak Ridge measuring the critical shear stress (minimum shearing stress required to produce plastic deformation in a single crystal) in pure copper. It increases rapidly with irradiation, see Fig. 9. The observed behaviour may probably be explained by pinning down of dislocations and of Frank-Read sources. The influence of temperature was also studied; critical shear stress in Cu as a function of temperature is shown in Fig. 10 for different values of the integrated neutron flux Φ.

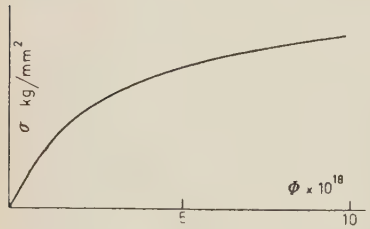


Fig. 9. – Critical shear stress of Cu as function of neutron flux.

Lattice expansion as observed by X-ray diffraction is in an approximately linear relation to the integrated flux, showing the influence of interstitial atoms.

A particularly striking effect was observed in AuCu which in ordered condition is tetragonal ($a > c$) and in complete disorder is cubic ($a = c$.) Bombardment modified the parameters in such a way as to bring them gradually to the same value: it is therefore possible to

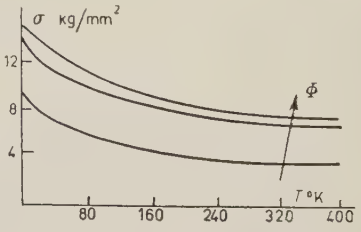


Fig. 10. – Critical shear stress of Cu as function of temperature at various neutron fluxes.

estimate in this way the value of the total flux that produces complete disorder, see Fig. 11.

Studying the *electric and magnetic properties* of Ni_3Mn disordered by fast neutrons, ARONIN found experimentally an average of 5 000 disordered atoms per collision. This result may be compared with the calculated 1 300 direct displacements and 6 100 replacements by the Kinchin and Pease mechanism, which amount to a total of 7 400 disordered atoms. It appears thus that in this case the Kinchin and Pease model is applicable. For gold-copper alloys the results are not so good as far as comparison with experiment is concerned.

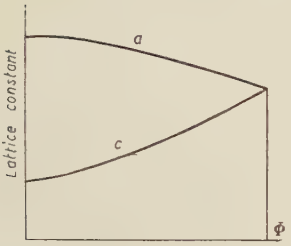


Fig. 11. – The lattice constants of AuCu as functions of neutron flux.

Phase change in the bombarded metal may be also observed; KONOBEEVSKY (Geneva Conference) first reported phase changes in a neutron bombarded uranium-molybdenum alloy. Similar experiments were also made by BLEIBERG and LUSTMAN who studied carefully the changes in electrical resistivity. It appears that a 9 % uranium-molybdenum alloy quenched from a high temperature retains the body-centered structure of the γ phase, which is stable at high temperature, see Fig. 12. Bombarding a mixture of α and ε phases, stable at room temperature, at the same composition, transforms it back to the γ phase. This is probably a consequence of melting and quenching which takes place in various spikes which are produced by high energy fission fragments.

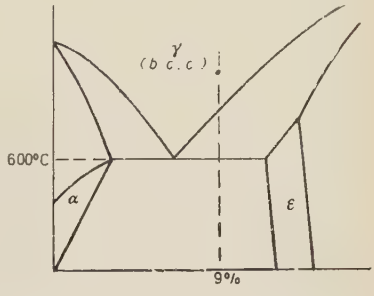
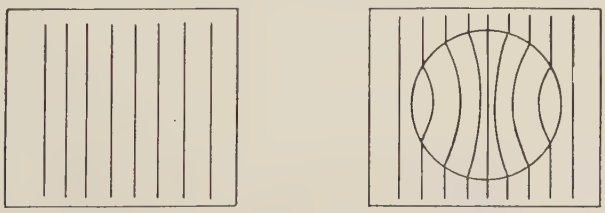


Fig. 12. – The diagram of uranium-molybdenum alloys.

Another evidence for thermal spikes has been obtained by GONSER and OKKERSE on single crystals of several compounds, such as GaSb and InSb, irradiated with deuterons. X-rays diffraction lines produced by the irradiated sections



(*a*) (*b*)

Fig. 13. - X-ray diffraction lines from InSb in an (*a*) unirradiated and (*b*) irradiated (in the central portion) condition.

of the crystal appear distorted and weak, while the lines from the non-irradiated region are straight and sharp (see Fig. 13). This effect is probably due to the local shrinking which occurs as a consequence of local melting in hot spots and subsequent quenching into an arrangement denser than the very open normal ordered crystal structure of these compounds.

20. – Alpha brass.

It is usually assumed that α brass is a random solid solution containing up to 33% Zn in Cu. A study was made of the change of its resistivity as measured at 80 °K on specimens which had been irradiated, for about a month, both at 80 °K and at room temperature in a pile. The results, in micro-ohm cm, are as follows:

% Zn	Low temperature irradiation	Room temperature irradiation
0	+ 0.037	—
10	+ 0.057	— 0.031
20	+ 0.019	— 0.062
30	+ 0.124	— 0.078

The most plausible interpretation of these results is that at low temperature, vacancies and interstitials produced on irradiation increase the resistivity by lowering the periodicity of the lattice. At room temperature, on the other hand, the mobility of defects introduced by irradiation accelerates the effective rate of diffusion, and the lattice approaches a state of short range order which exists at these temperatures. This lowers the resistivity. Measurements at 4 °K confirm the change of the residual resistivity. One would expect that for a long irradiation the increasing concentration of defects would eventually increase the resistivity even at room temperature.

KEATING, using neutron diffraction has checked that the long range order parameter in irradiated 30% brass is $S < 0.03$, but there is a possibility of the existence of short range order, corresponding to a quite low critical temperature.

21. – Alkali halides.

Lithium fluoride, exposed to a neutron flux, undergoes an internal bombardment due to the particles released in the reaction

$$^{6}\mathrm{Li}(\mathrm{n}, \alpha)\,^{3}\mathrm{H}$$

the energy of each particle being of the order of 2 MeV. The change of lattice constants as observed by X-ray diffraction is in close agreement with bulk dilatation. It may be concluded that the formation of Frenkel defects is the main consequence of irradiation. With increasing integrated neutron flux, the lattice expansion reaches a maximum (PERIO) after which, at still higher fluxes, it drops down nearly to zero (Fig. 14). A possible explanation of this reversal may be found in the coalescence of defects in platelets, which join the lattice so that the stresses are relieved. Small angle X-ray scattering and streaks in Laue diagram confirm the existence of clusters of vacancies and of flat defects presumably made up of interstitials (GUINIER).

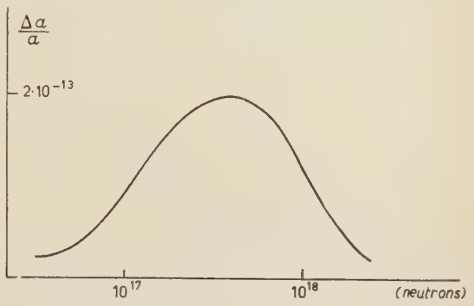


Fig. 14. – Change of lattice constant of LiF as function of neutron flux.

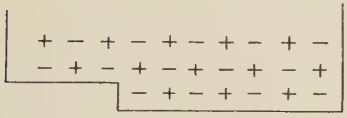


Fig. 15. – Jog in an incomplete plane of ions of an edge dislocation.

The production of defects in alkali halides by ionizing radiation may be explained by two mechanisms: « evaporation » of vacancies from dislocation jogs (SEITZ), where an equal number of positive and negative vacancies are produced (see Fig. 15), and the mechanism suggested by VARLEY. In this case a multiple ionization of a negative ion may leave it with a positive charge, so that it finds itself in a configuration of high energy, and can be easily shifted into an interstitial position by thermal agitation. In this way only negative vacancies

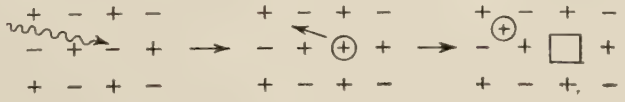


Fig. 16. – Varley mechanism of defect formation.

could be produced (Fig. 16). Actually the lifetime of the positive halogen ion may be too short to allow the occurrence of this phenomenon.

Effects of proton irradiation on alkali halides were observed by KOBAYASHI by studying changes in several different properties. The ionic conductivity of crystals irradiated at room temperature was measured as a function of gradually increasing temperature. Contrary to a naive expectation it increases by a factor up to 100 and then returns to normal values (Fig. 17). The release of stored energy shows a peak at about 200 °C, which coincides

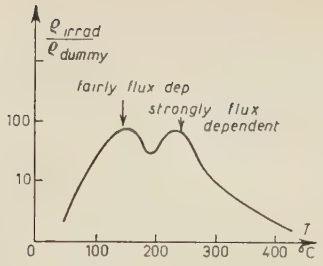


Fig. 17. – Ratio of resistivities of irradiated and non-irradiated NaCl as a function of gradually rising temperature.

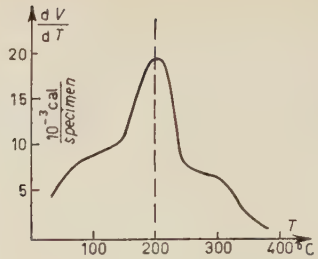


Fig. 18. – Stored energy released upon gradual heating of irradiated NaCl.

with the first resistivity drop (Fig. 18). From the total energy thus released one obtains the number of defects produced per incident proton, which is of the order 2000 to 3000.

Absorption spectra give evidence of the production of F, M and V_3 centres upon irradiation. Assuming the usual oscillator strengths, it is possible to evaluate the number of defects, which in a particular case was found to be

3200 negative ion vacancies
and 1300 positive ion vacancies
that is 4500 defects formed per incident proton per cm path.

The above results may be also correlated with density measurements which, using a flotation method, can be carried out with an accuracy of $2 \cdot 10^{-6}$. Here again one finds 2000 to 3000 defects per proton. With increasing temperature the density change behaves as shown in Fig. 19, which indicates that while, up to 200 °C, no defects were annealed out, many rearrangements took place. In particular, clusters (centres M and R') are formed. This would account for the increase of resistivity on annealing, since there are more negative than positive ion vacancies present, and the clustering eliminates the current carriers (i.e. positive ion vacancies) present in the crystal before irradiation. It seems thus that in order to account for the effects here

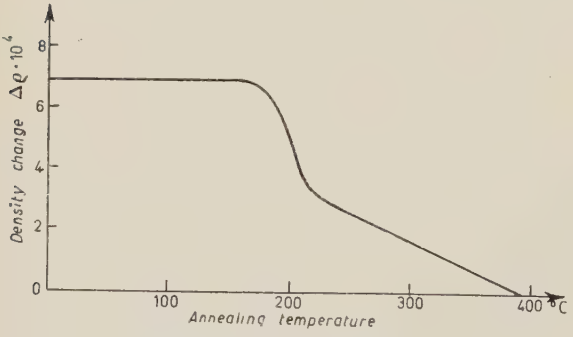


Fig. 19. – Annealing of density change of an irradiated NaCl crystal.

observed one needs both the « jog » mechanism and the Varley mechanism to be operative.

A change of elastic constants of KCl was observed with high energy proton irradiation, C_{ij} being reduced by 1 to 2%. Conclusive evidence has however not yet been obtained on this point.

22. – Semiconductors.

The threshold energy for displacement in germanium was determined first by KLONTZ using electron bombardment at low temperature and measuring the change in conductivity. He found

$$E_d \cong 31 \text{ eV} .$$

LOFERSKI and RAPPAPORT, on the other hand, measured the lifetime of minority carriers at n-p junctions. As a result of their experiments a considerably lower displacement energy

$$E_d = 13.9 \text{ eV}$$

was obtained for the same material.

Theoretical calculations of this value by KOHN give the result

$$E_d = 7 \text{ to } 15 \text{ eV}$$

for an atom pushed through a triangle of nearest neighbours. Displacement in the cubic direction may require

$$E_d = 20 \text{ to } 30 \text{ eV} .$$

In interpreting these results it should be remembered that the lifetime of the displaced atom may be different in the various interstitial positions, and there is a possibility that some of them may not exist long enough to affect the resistivity. Furthermore, different types of defects could be responsible for the change of resistivity and for the variation of lifetime. More recent measurements of resistivity made at Purdue University lowered the observed displacement threshold to

$$E_d = 17 \text{ eV} .$$

According to a more refined theory the probability of displacement should be assumed to be zero at E_d, and to rise with increasing energy reaching unity at an energy E'. This may account for the long tail in the experimental curve (Fig. 20).

Two donor and two acceptor levels are introduced in bombarded germanium; the energy level scheme may be pictured as in Fig. 21. The donors are associated with interstitials, the acceptors with vacancies.

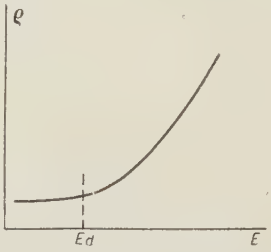


Fig. 20. – Resistivity of germanium as function of energy of incident electrons.

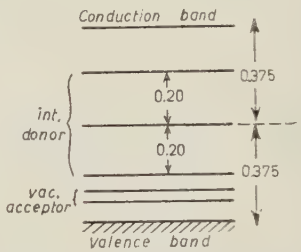


Fig. 21. – Donor and acceptor states introduced in germanium by irradiation.

Conductivity of an n-type Ge is first reduced, then after a sufficiently long bombardment the crystal shows an increasing p-type conductivity. At the minimum of the curve the number of defects produced just about balances the number of electrons originally present. The number of carriers lost per defect, as may be deduced from the initial slope of the curve, is roughly 1 to 2 (Fig. 22). The behaviour of p-type semiconductors under bombardment varies considerably. An example of the curve of conductivity as a function of total flux is also given in Fig. 22.

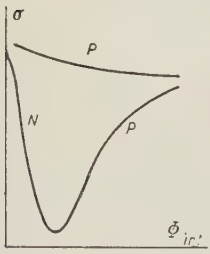


Fig. 22. – Conductivity of n and p-type germanium as a function of irradiation flux.

An indium atom, in Indium Antimonide bombarded by thermal neutrons, may undergo a transmutation into tin which is an impurity acting as donor. Shielding with cadmium cuts out this effect, as expected.

23. – Graphite.

Neutron transmission measurements on graphite made by DIENES and co-workers using long wave-length neutrons made possible an accurate evaluation of the total number of defects since they act as well known scattering centres. One obtains a concentration of displacements of 2.6 % for an integrated damaging neutron flux equal to $1.1 \cdot 10^{20}$ n/cm^2. The discrepancy with theory, by a factor of 2, is definitely lower than the discrepancy by a factor of 5 found in resistivity measurements and may be due to the lack of knowledge of the exact neutron flux and its spectrum.

Stored energy release as a function of the temperature of annealing is

given in Fig. 23; different curves correspond to different integrated neutron fluxes. It may be observed that with increasing flux the energy reaches a maximum at about 200 °C and then decreases. An interpretation of this phenomenon may be based on the assumption that interstitials produced by heavy irradiation, coalesce to form large stable clusters, not easily annealed. At low fluxes defects are distant and anneal out individually. Complete annealing of heavily irradiated graphite is only possible well above 1 000 °C.

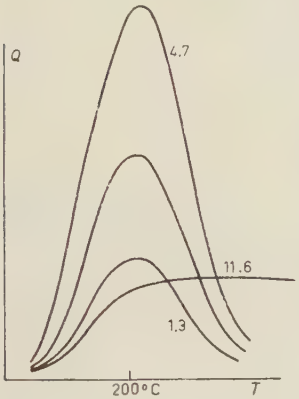


Fig. 23. – Release of stored energy from irradiated graphite as function of temperature and increasing flux.

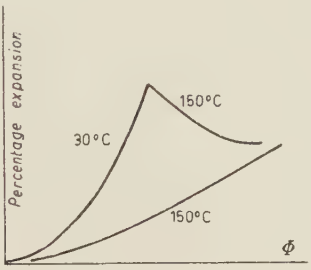


Fig. 24. – Expansion of graphite as function of neutron flux at various temperatures.

Measurements of percent expansion as a function of integrated neutron flux were made both at 150 °C and at 30 °C. As is shown in Fig. 24 the expansion at low temperature is bigger, but if after irradiation at 30 °C the temperature is raised to 150 °C the curve falls, giving a striking evidence of annealing during irradiation.

24. – Uranium.

Neutron irradiation of α-Uranium single crystals (orthorhombic) produces an increase of the lattice parameter in the (010) and a decrease in the (100) direction. This effect is very important in practice; its mechanism is not yet completely understood, but could be related to the formation of additional linear chains of atoms produced by interstitials while the vacancies migrate along the (100) direction until they disappear at the surface.

REFERENCES

[1] F. Seitz: *Disc. Faraday Soc.*, **5**, 271 (1949).
[2] G. H. Kinchin and R. S. Pease: *Reports on Progress in Physics*, **18**, 1 (1955).
[3] F. Seitz and J. S. Koehler: *Solid State Physics*, **2**, 305 (1956).
[4] H. Brooks: *Ann. Rev. Nucl. Sc.*, **6**, 215 (1956).
[5] G. J. Dienes and G. H. Vineyard: *Radiation Effects in Solids* (New York, 1957).
[6] J. W. Glen: *Adv. in Phys.*, **4**, 381 (1955).
[7] K. Lintner and E. Schmid: *Ergebn. Exakt. Naturw.*, **28**, 302 (1955).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Anelasticity of Metals.

C. ZENER

Westinghouse Research Laboratories - Pittsburgh, Pennsilvania

1. – Formal theory.

For the conventional formulation of Hooke's law the relation between an applied force F and the resulting displacement D is regarded as both linear and simultaneous. Thus

$$D(t) \sim F(t) .$$

More precisely,

$$D(t) = CF(t) , \tag{1a}$$

where C is the appropriate compliance constant. In the development of Hooke's law three generalizations have been made.

Firstly, Eq. (1a) may be generalized into a tensor equation relating the strain tensor to a stress tensor. We are thereby led to the conventional theory of elasticity.

Secondly, Eq. (1a) may be generalized by relaxing the linearity condition. We are thereby led to the theory of non-linear elasticity.

Thirdly, we may generalize Eq. (1a) by relaxing the simultaneity condition. Thus we may generalize (1a) into

$$D(t) = C'F(t) + \int_{-\infty}^{t} G(t-t')F(t')\,\mathrm{d}t' . \tag{1b}$$

This generalization leads to the theory of anelasticity, the subject of these lectures.

A material which obeys (1b) rather than (1a) is said to be « anelastic ».

It behaves as if it had a memory. Such behavior is in fact merely a manifestation of the existence of certain hidden parameters. As an example, suppose the relation between force and displacement is given by the following set of instantaneous equations with the hidden parameter P:

$$\left\{\begin{aligned} D(t) &= CF(t) + \alpha P(t)\,, \\ \dot{P}(t) &= -\tau^{-1}P(t) - \lambda F(t)\,. \end{aligned}\right. \tag{2}$$

The solution of this set of simultaneous equations leads to

$$D(t) = CF(t) - \alpha\lambda \int_{-\infty}^{t} \exp\left[-\frac{(t-t')}{\tau}\right] \dot{F}(t')\,\mathrm{d}t'\,. \tag{3}$$

This solution contains no evidence of the hidden parameter P. This simplification has however been purchased by the loss of simultaneity in the relation between force and displacement. Eq. (3) may finally be put into the form of Eq. (1*b*) by integration by parts.

$$D(t) = (C - \alpha\lambda)F(t) + \frac{\alpha\lambda}{\tau}\int_{\infty}^{t} \exp\left[-\frac{(t-t')}{\tau}\right] F(t')\,\mathrm{d}t'\,. \tag{4}$$

We shall find that the memory function may in fact be interpreted in terms of hidden parameters. The discovery and the investigation of these hidden parameters is in fact the justification for the study of the field of anelasticity. The precise differential equations which govern the behavior of the hidden parameters vary from case to case. These differential equations determine, of course, the precise dependence of the memory function upon its argument $t - t'$. In this lecture we shall study those properties of a material which are dependent only upon the general characteristics of this function.

Towards this end we shall introduce the normalized memory function $B(t - t')$,

$$B(s) = G(s) \Big/ \int_{0}^{\infty} G(s)\,\mathrm{d}s\,,$$

having the property

$$\int_{0}^{\infty} B(s)\,\mathrm{d}s = 1\,.$$

Thus in our above sample

$$B(s) = \tau^{-1} \exp\left[-s/\tau\right] .$$

We then rewrite Eq. (1*b*) as

$$D(t) = C'F(t) + C'' \int_{-\infty}^{t} B(t-t')F(t')\,\mathrm{d}t' . \tag{1c}$$

As a reflection of the fact that the effect of a force applied at $t = 0$ for a short interval $\mathrm{d}t$ must gradually become smaller and smaller with time, the unit memory function must be a monotonically decreasing function of time. Such a time dependence is illustrated in Fig. 1.

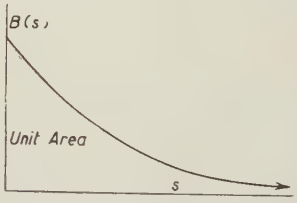


Fig. 1. – Time dependence of the unit memory function.

Now suppose that an F is suddenly applied and maintained constant at this value. The deformation thus suddenly jumps to the value $C'F$. C' is then the unrelaxed compliance constant. It will hereafter be referred to as C_u. After the instantaneous rise to C_uF, the deformation will gradually « creep » to the asymptotic value of $(C' + C'')F$. The quantity $C' + C''$ is called the relaxed compliance constant, and will be given the notation C_R. Thus Eq. (1*c*) may be written as

$$D(t) = C_uF(t) + (C_R - C_u) \int_{-\infty}^{t} B(t-t')F(t')\,\mathrm{d}t' . \tag{1d}$$

The phenomenon of creep is illustrated in Fig. 2.

A creep experiment provides us with a method not only of determining the two constants C_u and C_R, but also the precise form of the unit memory function $B(t)$. Thus upon differentiating Eq. (1*d*) under the conditions of a creep experiment, we find

$$\dot{D}(t) = (C_R - C_u)FB(t) .$$

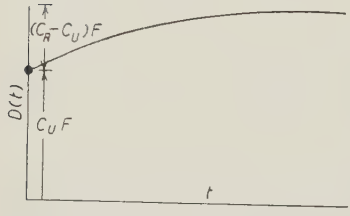


Fig. 2. – Time dependence of deformation in creep.

Now let us suppose that after the deformation has attained its asymptotic value the force is suddenly removed. The deformation then suddenly drops by the amount C_uF, and thereafter gradually relaxes towards zero. Upon taking the origin of time as the time the force

is removed, we obtain from Eq. (1*d*),

$$\dot{D}(t) = -(C_R - C_u) F B(t) .$$

This « creep recovery » gives us a second experimental method of determining the remembrance function. From the experimental point of view this method is preferable in that it is easier to maintain the force constant at the value zero than at any finite value. Creep recovery is illustrated in Fig. 3.

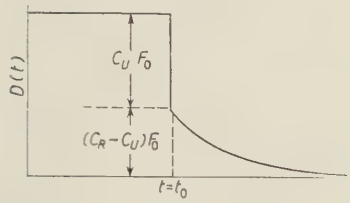


Fig. 3. – Creep recovery.

From the concept of creep, and of creep recovery, we arrive at the concept of hysteresis. Suppose we perform the following experiment,

t_0 to t_1: rapidly apply F,
t_1 to t_2: let F remain constant,
t_2 to t_3: rapidly remove F.

The D-F plot will then have the general form of Fig. 4. The area of the closed loop is just the work performed upon the system by external forces. From the second law of thermodynamics this work must be positive. The specimen cannot perform work upon the external world. The memory function G must therefore be positive. From inspection of Eq. (4) we therefore conclude that the product $\alpha\lambda$ of the constants in Eq. (3) must be positive. Likewise the constant C'' in (5) must also be positive.

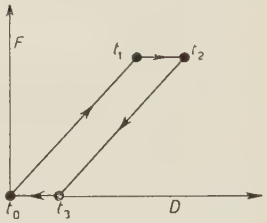


Fig. 4. – Force-deformation relation in mechanical hysteresis.

The positive character of the memory function may be regarded as a reflection of Chatelier's principle. Thus suppose we suddenly increase the displacement from 0 to a constant value D_0. The force is then given by the equation

$$C_u F(t) = D_0 - \int_0^t G(t - t') F(t') \, dt' , \qquad t > 0.$$

The readjustment of the hidden parameters thus reduces F, i.e. opposes the original force.

Hysteresis is commonly measured under conditions of periodic variations in force. In order to analyze this situation, we rewrite Eq. (1*d*) as

$$D(t) = C_u F(t) + (C_R - C_u) \int_0^\infty B(t') F(t - t') \, dt' . \tag{1e}$$

Upon replacing $F(t)$ by $F_0 \cos \omega t$, we obtain

$$(1f) \qquad D(t) = \{C_u + (C_R - C_u)\, C(\omega)\} F_0 \cos \omega t + (C_R - C_u) S(\omega) F_0 \sin \omega t,$$

where $C(\omega)$ and $S(\omega)$ are the cosine and sine transforms of $B(t)$, respectively. This solution is plotted in Fig. 5. The parallelogram gives D vs. F for the case where F suddenly reverses between $\pm F_0$, the time between each reversal being sufficiently long as to obtain complete relaxation. The ellipse gives D vs. F for the case where F varies sinusoidally with time.

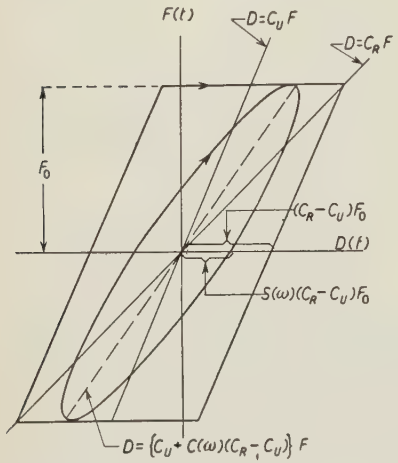


Fig. 5. – Special cases of mechanical-hysteresis loops.

The area of a hysteresis loop is given by

$$\text{Area} = \nu^{-1} \overline{\dot{D}F} = \pi (C_R - C_u) S(\omega) F_0^2 .$$

We are usually more interested in a dimensionless measure of the hysteresis, Such a dimensionless measure is provided by $\operatorname{tg} \alpha$, where α is the angle by which deformation lags behind the applied force. In deference to electrical engineers, we shall denote $\operatorname{tg} \alpha$ by Q^{-1}. When Q^{-1} is very small compared to unity, we have from $(1f)$

$$Q^{-1} = \Delta_c S(\omega) ,$$

where

$$\Delta_c = (C_R - C_u)/C_u$$

is known as the relaxation strength referred to the compliance constant. Under the conditions of small Q^{-1} this quantity may likewise be interpreted as

$$(5) \qquad Q^{-1} = \frac{\text{Area of hysteresis loop}}{2\pi \text{ maximum strain energy}} .$$

Under these same conditions,

$$Q^{-1} = \pi^{-1} \text{ logarithmic decrement.}$$

In the literature $Q^{-1}(\omega)$ is referred to as the « internal friction ».

Simplified expressions may be obtained for $Q^{-1}(\omega)$ in the two extreme cases

$$\omega\tau \ll 1 \qquad \text{and} \qquad \omega\tau \gg 1 ,$$

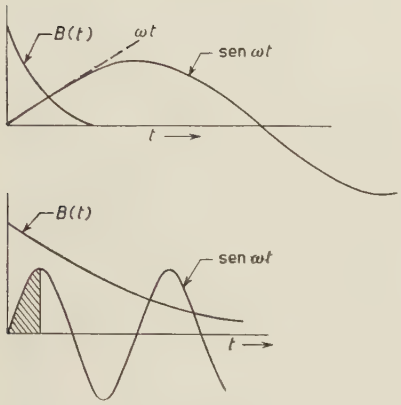


Fig. 6.

where the relaxation time τ is given by

$$\tau = \int_0^\infty tB(t)\,\mathrm{d}t . \tag{6}$$

Thus, referring to Fig. 6

$$S(\omega) \cong \int_0^\infty \omega t B(t)\,\mathrm{d}t = \omega\tau , \qquad \omega\tau \ll 1 \tag{7}$$

and

$$S(\omega) \cong \int_0^{\omega t = \pi/2} \sin \omega t B(0)\,\mathrm{d}t = \frac{B(0)}{\omega} , \quad \omega\tau \gg 1. \tag{8}$$

In addition, the following integral relation is obtained,

$$\int_0^\infty \frac{S(\omega)}{\omega}\,\mathrm{d}\omega = \int_0^\infty \int_0^\infty B(t)\,\frac{\sin \omega t}{\omega}\,\mathrm{d}t\,\mathrm{d}\omega = \frac{\pi}{2}\int_0^\infty B(t)\,\mathrm{d}t ,$$

and hence

$$\int_0^\infty S(\omega)\,\mathrm{d}\ln\omega = \frac{\pi}{2} . \tag{9}$$

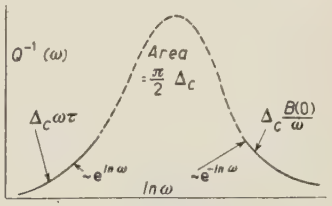


Fig. 7. – Frequency dependence of the internal friction.

These relations are represented in Fig. 7.

In the particular case that our hidden parameter is governed by Eq. (2),

$$C_R - C_u = \alpha\lambda , \tag{10}$$

$$B(t) = \exp\left[-t/\tau\right]/\tau \tag{11}$$

and

$$C(\omega) = \frac{1}{1 + (\omega\tau)^2} , \tag{12}$$

$$S(\omega) = \frac{\omega\tau}{1 + (\omega\tau)^2} . \tag{13}$$

2. – Thermal relaxation.

The first hidden parameter we shall consider is temperature. The analysis of this parameter requires no new concepts, and its effects have been subjected to rigid experimental tests.

If the changes in stress within a material occurred so slowly that the temperature remained constant throughout, Hooke's law would prevail, and the isothermal compliance constant would be appropriate. On the other hand, if the changes in stress were so rapid that no appreciable heat flow took place, Hooke's law would again be valid and the adiabatic compliance functions would be appropriate. In the general case changes in stress cannot be regarded as taking place either isothermally or adiabatically, and the anelastic equation (1*d*) must be used. As in all anelastic materials, two quantities are of importance, the relaxation strength, and the remembrance function or one of its transforms. We shall first consider the relaxation strength.

Let X_T and X_S be the isothermal and adiabatic compressibilities, respectively, corresponding to the relaxed and unrelaxed compliance constants of the previous chapter. We are to evaluate the relaxation strength

$$\Delta_X = (X_T - X_S)/X_S \,.$$

Towards this end we write the set of equations

$$\mathrm{d}v = X_T(-\mathrm{d}P) + \beta\,\mathrm{d}T \,,$$

$$\mathrm{d}S = \beta(-\mathrm{d}P) + (C/T)\,\mathrm{d}T \,,$$

where $\mathrm{d}v$ and $\mathrm{d}S$ refer to an element having a unit volume under standard conditions. Here β is the volume thermal expansion coefficient, C is the thermal capacity per unit volume. The identity of the cross coefficients is insured by the fact that

$$\mathrm{d}(-vP + ST - \varepsilon) = v(-\mathrm{d}P) + S\,\mathrm{d}T$$

is a perfect differential.

We thereby obtain

$$X_S = X_T - T\beta^2/C \,,$$

and hence

$$\Delta_X = T\beta^2/CX_S \,. \tag{1}$$

For a monatomic gas one deduces

$$\Delta_x = \tfrac{2}{3} .$$

In order that we may obtain a physical feeling for the order of magnitude of Δ_x for metals we shall write the above equation as

$$\Delta_x = (\beta T_m)(\beta / C X_s)(T/T_m) .$$

Now all close packed metals have essentially the same value for the first factor, namely

$$\beta T_m = 0.06 \pm 0.01 .$$

The second dimensionless factor occurs frequently in physics. It is called Grüneisen's constant, and is given the symbol γ. The physical interpretation of γ is obtained by considering that thermal expansion is generated by an internal pressure arising from the thermal energy. According to this view,

$$\beta = \frac{1}{v} \frac{\partial v}{\partial P_T} (\delta P_{\text{int.}}/\delta U)(\delta U/\delta T) .$$

We therefore identify

$$\gamma = \delta P_{\text{interval}}/\delta U .$$

Again, for close packed cubic metals, γ clusters about a common value:

$$\gamma = 2.95 \pm 0.08 .$$

(Ca is an exception.)

For these metals we therefore obtain

$$\Delta_x \cong 0.18(T/T_m) .$$

The coefficient is somewhat smaller for b.c.c. metals.

The relaxation strength for thermal relaxation in uniaxial tension is obtained from Eq. (2.1) upon replacing β by the linear expansion coefficient α, and X_s by the compliance constant for uniaxial tension. Now the bulk modulus K and Young's modulus Y are nearly identical, being precisely identical for a Poisson ratio of $\frac{1}{3}$. We thereby obtain

$$\Delta_Y \cong 0.02(T/T_m) .$$

We next proceed to obtain the remembrance function for a particular case.

The case we shall choose is that of a reed vibrating transversely. It is more convenient in this case to regard the tensile strain ε as the independent variable rather than the stress σ. The governing equations for this system are

$$\text{(2)} \qquad \begin{cases} \sigma = Y\varepsilon - \lambda \Delta T\,, \\ \dot{\Delta T} = D\nabla^2 \Delta T - \mu\varepsilon\,, \\ \varepsilon = \varepsilon_0 (x/\tfrac{1}{2}d)\cos \omega t\,. \end{cases}$$

Here ΔT denotes the fluctuation in temperature from its mean value, D is the thermal diffusion coefficient, d is the thickness of the reed, and ω is the angular frequency of vibration.

In order [1] to evaluate $Q^{-1}(\omega)$ we introduce the normalized characteristic solution $U_k(X)$ of

$$D\nabla^2 \Delta T = \tau^{-1} \Delta T\,,$$
$$\text{grad}_m\, \Delta T = 0 \quad \text{at surface.}$$

The boundary condition corresponds to a zero heat flux across the surface. We then expand

$$\varepsilon(x,t) = \sum_k \varepsilon_k(t)\, U_k(x)\,,$$
$$\sigma(x,t) = \sum_k \sigma_k(t)\, U_k(x)\,,$$
$$\Delta T(x,t) = \sum_k \Delta T_k(t)\, U_k(x)\,.$$

Equation (2) then becomes

$$\sigma_k = Y\varepsilon_k - \lambda\, \Delta T_k\,,$$
$$\Delta \dot{T}_k = -\tau_k^{-1}\, \Delta T_k - \mu \dot{\varepsilon}_k\,,$$
$$\varepsilon_k = \varepsilon_k(t)\,, \quad \text{known function.}$$

Using the definition for $Q^{-1}(\omega)$ given by 1.5, we obtain

$$\text{(3)} \qquad Q^{-1}(\omega) = \Delta_Y \sum_k f_k S_k(\omega)\,.$$

Here f_k is a weighting function having the property

$$\sum f_k = 1\,.$$

Specifically

$$f_k = 96\pi^{-4}(2k+1)^{-4},$$

$$f_0 = 0.986, \qquad f_1 = 0.012, \qquad f_2 = 0.0016.$$

Due to the very rapid convergence of the f's, only a negligible error would be introduced by neglecting all weighting functions other than the first. One thereby obtains

$$Q^{-1} = \Delta_r \cdot \frac{\omega\tau}{1+(\omega\tau)^2}. \tag{4}$$

with

$$\tau = (d^2/\pi^2 D).$$

A careful experimental test of Eq. (4), or more accurately of Eq. (3), is given in Fig. 8 after BERRY [2]. The solid curve is the theoretical equation (3) with no adjustable parameters.

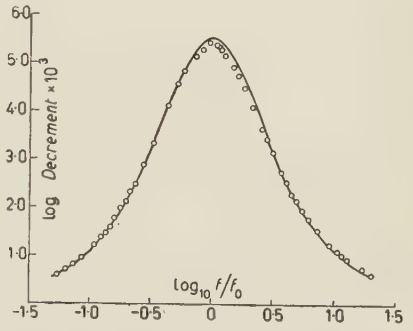


Fig. 8. – Frequency-dependence of internal friction of four annealed α-brass reeds. Actual frequencies f are divided by the frequency of peak damping f_0. The curve is given by theory with no arbitrary parameters. After BERRY.

The elastic anisotropy of the individual crystallites in a single specimen gives rise to microscopic thermal currents between adjacent crystallites [3]. Again a straightforward calculation may be made for the time of relaxation as a function of grain size G and thermal diffusion coefficient. Thus a spherical crystallite of radius a, subject to a constant temperature on the surface, has a thermal relaxation time given by

$$\tau = a^2/(\pi^2 D).$$

But the customarily measured grain size G is related to a by

$$G^2 = \pi \int_0^a (a^2 - x^2)\,\mathrm{d}x \Big/ \int_0^a \mathrm{d}x = (2/3)\pi a^2.$$

A maximum internal friction would therefore be obtained at a frequency $\nu = 1/2\pi\tau$, i.e., at

$$\nu G^2/D = \pi^2/3 = 3.3.$$

The observed maximum is indeed found [3], within experimental error, to occur for just this value of the dimensionless parameter $\nu a^2/D$.

Our physical insight as to the mechanics of damping through thermal flow may be furthered by a closer examination of our basic eq. (1.2). For this examination we shall regard the stress P and temperature fluctuation ΔT as the independent variables. Thus

$$\varepsilon = C_R\sigma + \alpha\Delta T\,,$$
$$\dot{\Delta T} = D\nabla^2\Delta T - \lambda\dot{\sigma}\,.$$

We now note that the rate of change of entropy density, $\dot{S}$, is given by

$$\dot{S} = (k/T)\nabla^2 T = (CD/T)\nabla^2 T\,,$$

where k is the thermal conductivity, C the specific heat per unit volume. The second of the above equations may thus be rewritten as

$$\dot{S} = (C\lambda/T)\dot{\sigma} + (C/T)\Delta T\,.$$

Our fundamental set of equations is therefore

$$\varepsilon = C_R\sigma + \alpha\Delta T\,,$$
$$\delta S = (C\lambda/T)\sigma + (C/T)\Delta T\,.$$

From the requirement that

$$\mathrm{d}E = \sigma\,\mathrm{d}\varepsilon + T\,\delta S\,,$$

be a perfect differential, we establish the identity of the cross coefficients.

$$(C\lambda/T) = \alpha\,.$$

Our fundamental equations thus assume the symmetrical form

$$\varepsilon = C_R\delta\sigma + \alpha\Delta T\,,$$
$$\delta S = \alpha\,\delta\sigma + (C/T)\Delta T\,.$$

Physicists are accustomed to two types of symmetrical relations. On the one hand, such relations occur between displacements X_i and the applied forces, f_i,

$$X_i = \sum_j C_{ij} f_j\,, \qquad C_{ij} = C_{ji}\,,$$

providing no irreversible processes are occurring. On the other hand, such relations occur between fluxes J_i, and the corresponding driving potentials φ_j,

$$J_i = \sum_j C_{ij}\varphi_j\,, \qquad C_{ij} = C_{ji}\,,$$

provided the potentials are so normalized that

$$\dot{S} = \sum_j J_j\,\varphi_j\,.$$

Our set of equations belongs to neither of these two classes. Our ε corresponds to a displacement rather than to a flux. However, the increment of δS during a complete cycle is itself a measure of the irreversibility of our processes.

Let us now look at the energetics of the dissipation of energy during one cycle of vibration. We shall consider the cycle as beginning at a time at which $\sigma = 0$, and as ending at the second time at which σ returns to zero. The work δW done upon an element of volume δv is then

$$\frac{\delta W}{\delta v} = \oint \sigma\,\mathrm{d}\varepsilon = \alpha \oint \sigma\,\delta T\,.$$

The net heat enflux δQ into the same element of volume is given by

$$\frac{\delta Q}{\delta v} = \oint T\,\mathrm{d}S = \alpha \oint T\,\delta\sigma + C\oint \delta T\,,$$

or $\delta Q = -\delta W + C\Delta T$, where ΔT is the net change in T during cycle. Now in the system we have considered, the heat flux across the boundaries is zero. Hence the integral of δQ over the entire specimen is precisely zero. Upon performing this integration we obtain

$$\text{net work performed} = \oint C\Delta T\,\mathrm{d}v\,.$$

Further insight may be gained by examining directly the entropy increase per cycle,

$$\Delta S - \oint \frac{\delta Q}{T}\,.$$

Set

$$T = T_0 + \Delta T\,.$$

Thus

$$\Delta S = -\,T_0^{\,2} \oint T\,\delta Q\,.$$

Now under cyclic conditions

$$\dot{Q} \sim -\Delta T$$

and hence the total heat which has flowed into the element, ΔQ, is 90° out of phase with ΔT. We thereby obtain

$$\Delta S = (\pi/T_0^2)\,\Delta T_{\max} \cdot \Delta Q_{\max} .$$

Now when

$$\nu \to 0 , \qquad \Delta S \to 0 \qquad \text{because } \Delta T_{\max} \to 0 .$$

On the other hand, when

$$\nu \to \infty , \qquad \Delta S \to 0 \qquad \text{because } \Delta Q_{\max} \to 0 .$$

3. – Interstitial atoms in body centered cubic lattices.

The most widely studied type of relaxation is that associated with interstitial solute atoms in body centered cubic (b.c.c.) lattices. The physical origin of this relaxation may be readily interpreted. The interstitial sites in b.c.c. lattices have tetragonal symmetry. They are located at the centers of cube edges, or at the crystallographically equivalent positions at the centers of the cube faces. These sites are of three types, corresponding to the three principal axes to which the tetragonal axes may be parallel. In the absence of an externally applied stress, the three types of positions are energetically equivalent. The equilibrium distribution of atoms among the three types of positions is thus a random distribution. When, however, a tensile stress is applied along one axis, say the x axis, the interstitial sites associated with this axis now have a lower potential energy than the sites associated with the other two axes. The equilibrium distribution is then one in which more interstitial atoms are in the preferred sites than correspond to a random distribution. The unrelaxed modulus E_u is observed under conditions in which the distribution does not have time to relax. The relaxed modulus E_R is observed under conditions in which the equilibrium distribution is maintained.

One reason for the exhaustive studies of this type of relaxation is the wealth of useful metallurgical information which such studies can give. Thus the relaxation strength Δ_E is proportional to that concentration of foreign atoms which are in solid solution, in contrast to the concentration of foreign atoms which are in some sort of precipitate. Through an appropriate study of Δ_E one may thus derive useful metallurgical information in regard to the

saturation concentration as a function of temperature, the mechanics of precipitation, etc. Again, the relaxation time τ for the establishment of equilibrium is essentially the mean time τ a foreign atom resides in a given interstitial site before jumping into an adjacent site. A study of the variation of τ with temperature thus leads directly to the temperature dependence of the diffusion coefficient.

A second reason for the exhaustive studies of this type of relaxation is that the very simplicity of the relaxation, combined with the ease of reproducing identical specimens, renders such experiments a convenient method of comparing with experiments the results of detailed theoretical deductions.

Particularly precise experimental work has been made on interstitial diffusion in b.c.c. lattices. Whereas essentially precise agreement has been obtained with theory, this theory was in fact based upon assumptions which have not heretofore been justified. In this chapter the theory for interstitial diffusion is examined once again, and it is found possible to base this theory upon firmer foundations.

Our first task will be to see if the relaxation of the distribution function is governed by a single relaxation time. Towards this end we denote by M_1, M_2 and M_3 the number per cm^3 of interstitial atoms in sites of type 1, 2 and 3, respectively. We denote by δM_1, δM_2, δM_3 the deviation of M_1, M_2, M_3 from values corresponding to a random distribution. Now every site of type 1 has two nearest neighbors of type 2 and two nearest neighbors of type 3. Corresponding relations hold for the other two types. If $\bar{\tau}$ denotes the mean time an interstitial atom remains in a particular site before jumping, then

$$\bar{\tau}\delta\dot{M}_1 = -\delta M_1 + (\tfrac{1}{2})(\delta M_2 + \delta M_3)\,,$$

with corresponding relations for δM_2 and δM_3. We now set

$$\delta M_j = A_j \exp[-t/\tau]$$

and solve the secular equation for τ. One root is ∞, and the corresponding solution for (A_1, A_2, A_3) is the trivial solution $(0, 0, 0)$. The other two roots of τ are identical, and have the value

$$\tau = (\tfrac{2}{3})\bar{\tau}\,. \tag{1}$$

Now τ may be determied from anelastic measurements. The mean time of stay $\bar{\tau}$ is then deducible from (1).

We next determine [4] the diffusion coefficient from the interstitial atoms in terms of $\bar{\tau}$. We consider diffusion occurring along the x axis, and number consecutive planes j, $j+1$, ... If $N_j(t)$ denotes the number of interstitial

atoms on plane j at time t, we find by crystallographic inspection

$$N_j(t+\bar{\tau}) = \tfrac{1}{6}\left\{N_{j-1}(t) + N_{j+1}(t)\right\} + \tfrac{2}{3}N_j(t)\,.$$

This equation is now rewritten as

$$N_j(t+\bar{\tau}) - N_j(t) = \tfrac{1}{6}\left\{N_{j+1}(t) - 2N_j(t) + N_{j-1}(t)\right\},$$

or, upon dividing by $\bar{\tau}$, and recognizing that the distance between planes is half a lattice constant,

$$\frac{\partial N}{\partial t} = D\,\frac{\partial^2 N}{\partial x^2}\,,$$

where

$$D = A^2/(24\bar{\tau})\,. \tag{2}$$

Here A is the lattice constant. We thereby obtain D in terms of the experimental time of relaxation

$$D = A^2/(36\tau)\,. \tag{2a}$$

Our next task will be the theoretical calculation of the mean time of stay $\bar{\tau}$. In this calculation we shall assume that the mass of the interstitial atoms is sufficiently large that quantum effects may be neglected. Now

$$1/\bar{\tau} = 4r\,, \tag{3}$$

where r is the initial rate at which an interstitial atom escapes into a particular neighboring site. This rate is given by

$$r = P_s \boldsymbol{V}\,,$$

where $P\,\delta x$ is the probability that the diffusing atoms be on the appropriate saddle position within the element of distance δx, and $\boldsymbol{V}$ is the mean forward velocity of the atoms in this element of distance. We have taken x as the co-ordinate joining the two interstitial sites, and s as the saddle position.

In a previous discussion of this problem, the author assumed without discussion that P_s could be represented by the Boltzmann distribution. This assumption leads to the equation

$$r = \nu \exp\left[-\Delta G/kT\right], \tag{4}$$

where ν is the frequency of vibration of the interstitial atom in its equilibrium position, and where the work function ΔG is defined as follows. We consider the interstitial atom to be confined to a vibration in a plane normal to the x axis. The ΔG is the work we perform in adiabatically moving the confining plane from a position passing through the equilibrium position to a position passing through the saddle point. Since

$$\left\{\begin{aligned} &\Delta G = \Delta H - T\Delta S \\ &r = \nu \exp[\Delta S/k] \exp[-\Delta H/kT] . \end{aligned}\right. \tag{5}$$

Since

$$\frac{\mathrm{d}(\Delta G/T)}{\mathrm{d}(1/T)} = \Delta H ,$$

irrespective as to whether ΔH is a function of T, the quantity ΔH, in eq. (5) is the heat of activation for diffusion. Upon combining (2)–(5) we obtain

$$D = D_0 \exp[-\Delta H/kT] , \tag{6}$$

with

$$D_0 = \tfrac{1}{6} A^2 \nu \exp[\Delta S/k] .$$

Since D_0 can be measured with considerable accuracy, it is worthwhile to estimate its value. The frequency ν may be estimated on the assumption that the work function G is a sinusoidal function of x. The entropy ΔS may be estimated on the assumption that the work ΔG goes primarily into straining the lattice, and that the temperature dependence of this work comes from the temperature dependence of the elastic modulus. Now D has been measured [5] for carbon in iron over 14 cycles of ten as illustrated in Fig. 9, resulting in a very accurate measurement of D_0. The D_0 obtained from the above assumption is in complete agreement [4] with that obtained experimentally.

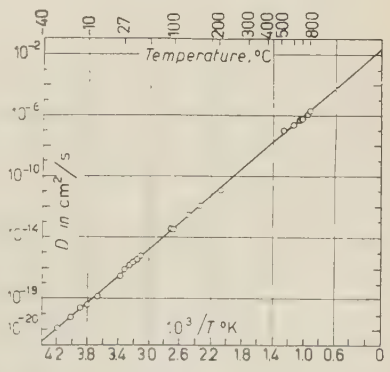


Fig. 9. Diffusion coefficient of carbon in α iron. (After WERT [5]).

In spite of the success of the above theoretical treatment for D_0, we must worry about the basic assumption leading to our expression for D_0, namely that the probability function P_s may be represented by an equilibrium Boltzmann function. One may

in fact reason against this assumption as follows. Just before our diffusing atom had sufficient energy to arrive at the saddle position, it had a somewhat lower energy E_1. Since every atom arriving at the saddle point passes out of the initial interstitial site, the distribution of atoms having the energy E_1 below the Boltzmann distribution will in turn result in a lowering of P_s below the Boltzmann value.

Reference to Fig. 10 will help clarify the above dilemma. A highly excited interstitial atom will have an accommodation coefficient α such that after each half cycle of vibration its energy is, on the average, equal to its original energy times α. A highly excited atom with energy E_1 will have a very great probability of losing an appreciable fraction of its energy during the next half cycle. It will have only a very slight probability of absorbing energy from the lattice, and thereby arriving at the saddle point. The slight drain of the population of the energy state E_1 via the saddle point will thus have only a negligible effect upon the population of the state E_1.

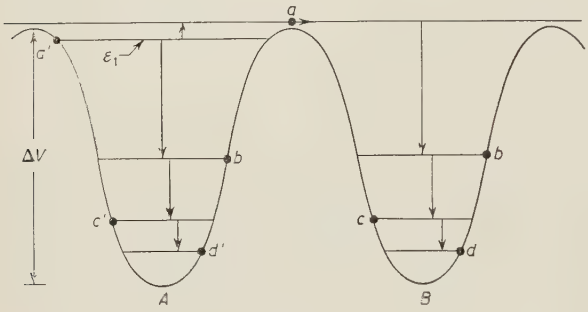


Fig. 10.

4. – Grain boundary relaxation.

Acting upon a suspicion that the grain boundaries of polycrystalline metals behave in a viscous-like manner, the author [6] estimated in 1941 the influence of shear stress relaxation across these boundaries upon the observed Young's modulus. Subsequent experiments on anelasticity of metals have verified the essential correctness of this assumption of viscous-like behavior of grain boundaries at least in those cases where the grain boundaries are free of slip inhibiting precipitate particles. In recent years this viscous slip across grain boundaries has been recognized as of particular practical importance at elevated temperatures in giving rise to stress concentrations at the junction of three grains, and thereby initiating fracture.

In order to render a polycrystalline system amenable to calculation, the author considered a typical grain to have a spherical shape. Even with this simplification the calculations were very lengthy. Now the existence of an Airy stress function for two dimensional systems greatly simplifies our problem

when we pass from a three to a two dimensional model of a polycrystalline aggregate. Because of the current use which is frequently made of the concept of viscous-like behavior of grain boundaries, it is believed appropriate to present the simplified calculation for the two dimensional model.

We consider a two dimensional polycrystalline aggregate subjected to an applied shear stress λ. The strain energy dnesity $\mathcal{E}$, and the shear modulus G, are related to one another through the relation

$$\mathcal{E} = \gamma^2/2G \,.$$

We now consider in detail two extreme cases. In the first case no shear stress relaxation takes place across the grain boundaries. Stress and strain are then uniform throughout the entire specimen, and consequently the energy density is uniform even on a micro-scale. The observed shear modulus is then identical with the shear modulus of the individual grain, and will be denoted by G_u. The second extreme case we shall consider is that in which complete shear stress relaxation has taken place across the grain boundaries. The stress system, and hence the strain energy density, will be highly inhomogeneous throughout the individual grains. It is the purpose of this section to calculate the average strain energy density in this case, and then to evaluate the relaxed shear modulus G_R through the relation

$$G_R = \gamma^2/2\mathcal{E} \,.$$

A typical grain will have the shape of a hexagon. In order to simplify calculations, we will replace this shape by a circle. The stress system we shall attempt to find is then that in which

$$\begin{aligned} \overline{X}_y &= \gamma \,, \\ \overline{X}_x &= \overline{Y}_y = 0 \,, \\ \widehat{\theta r} &= 0 \quad \text{at} \quad r = A \,, \end{aligned}$$

where A is the radius of the grain under consideration. We start our search by assuming an Airy stress function corresponding to our unrelaxed stress system:

$$\chi_1 = -\gamma xy = -(\gamma/2) r^2 \sin 2\theta \,.$$

Upon utilizing the relation

$$X_x = \frac{\partial \chi^2}{\partial y^2} \,, \qquad Y_y = \frac{\partial^2 \chi}{\partial x^2} \,, \qquad X_y = -\frac{\partial^2 \chi}{\partial x \, \partial y} \,,$$

$$\widehat{\theta r} = \frac{\partial}{\partial r} r^{-1} \frac{\partial}{\partial \theta} \chi \,,$$

we see that χ_1 satisfies the first two conditions but not the third. In fact

$$\widehat{\theta r} = \gamma \cos 2\theta .$$

We now add to χ_1 an Airy function χ_2 which just cancels the above $\widehat{\theta r}$, at the same time satisfying the condition that $\overline{X}_x$ and $\overline{Y}_y$ be zero. Such an Airy function is

$$\chi_2 = (\gamma/3A^2)(x^3y + xy^3) = (\gamma/6A^2)r^4 \sin 2\theta .$$

Now χ_2 gives a shear stress Y_y whose average is $-\gamma/2$. The Airy function which satisfies all our conditions is thus

$$\chi = 2(\chi_1 + \chi_2) .$$

A straightforward calculation now gives the average strain energy density of the grain as

$$\bar{\mathcal{E}} = (\gamma^2/2)\{(8/27)K^{-1} + (14/9)G_u^{-1}\} = (\gamma^2/2)G_R^{-1} .$$

From the relation

$$K^{-1} = \frac{3(1 - 2\sigma)}{2(1 + \sigma)} G_u^{-1} ,$$

where σ is Poisson's ratio, we obtain

$$G_R/G_u = \tfrac{3}{2}(1 + \sigma)/(3 + \sigma) .$$

As σ ranges over the physically permissible range from 0 to 0.5, the ratio G_R/G_u changes from 0.5 to 0.644. In particular, for the typical case of $\sigma = \frac{1}{3}$,

$$G_R/G_u = 0.60 , \qquad (\sigma = \tfrac{1}{3}) .$$

The experimental values of this ratio are given in the following table.

Metal	G_R/G_u	References
Al	0.66	[7]
Fe	0.70	[8]
Brass	0.70	[9]

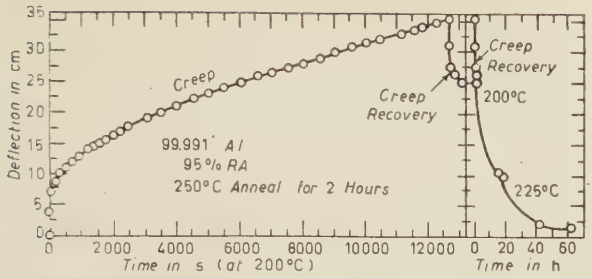


Fig. 11. – Creep and creep recovery in cold-worked aluminum presented on a linear scale. (After Kê [5]).

A relaxation closely related to grain boundary relaxation, but not yet completely understood, occurs in cold worked metals [10]. As illustrated in Fig. 11 and 12 when a cold worked specimen is annealed just below the recrystallization temperature, very large anelastic effects are manifested. Relaxed moduli, only one ninth of the unrelaxed moduli were found. More recent experiments [11] from Crussard's laboratory confirm these results, and demonstrate that the observed relaxation is indeed associated with a polygonized structure.

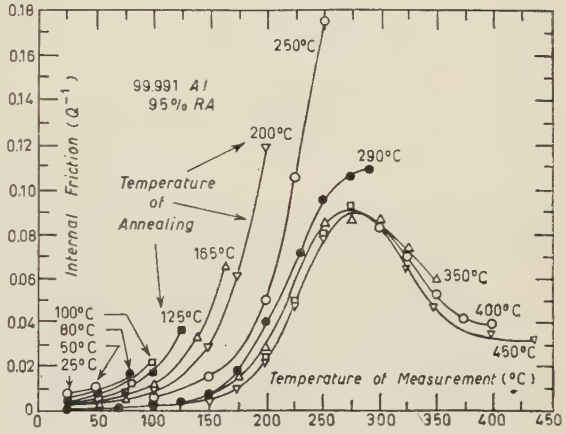


Fig. 12. – Influence of temperature of measurement and temperature of anneal upon internal friction of previously cold-worked aluminum. (After Kê [3]).

5. – Electron relaxation.

An elastic strain changes the equilibrium distribution of electrons in k space. The attainment of equilibrium will be governed by a relaxation time, τ_R. The equilibrium distribution may be changed in several ways, each way having a characteristic τ_R. If the angular frequency of vibration ω is such that $\tau_R\omega \ll 1$, the equilibrium distribution will be essentially maintained. In the other extreme case of $\tau_R\omega \gg 1$, no essential changes in the distribution will take place. Irreversible effects will be largest in the intermediate range $\tau_R\omega \sim 1$.

The first type of relaxation we shall consider is that associated with the distortion of the Fermi surface by strain. Thus a tensile strain would change a spherical surface into an ellipse. The first question to be answered is under what conditions can we obtain $\tau_R\omega \sim 1$. Under conditions usually encountered, τ_R is orders of magnitude smaller than the period of oscillation. Now the product $\tau_R\omega$ can be raised both by going to lower temperatures, and thereby increasing τ_R, and by going to higher frequencies. Both methods raise the

ratio l_m/λ, where l_m is the mean free path of the electrons with respect to momentum transfer, λ is the wave length of the elastic wave. In order for relaxation to occur, it is necessary that this ratio be less than unity. We must thus satisfy the two conditions:

(1) $$\tau_R \omega = 1 ,$$

(2) $$l_m \ll \lambda .$$

Upon dividing the second condition by the first, we obtain the condition

(3) $$\frac{\tau_m}{\tau_R} \ll 2\pi \frac{C}{v} ,$$

where v is the mean velocity of the electron, C is the velocity of sound.

If we have only a single Fermi surface,

$$\tau_R \cong \tau_m ,$$

and (3) becomes

(4) $$2\pi C/v \gg 1 .$$

Now for electrons

$$v = 60 \cdot 10^6 \ E^{\frac{1}{2}} \text{ cm/s} ,$$

where E is the electron energy in electronvolts, while for sound waves

$$C \cong 0.6 \cdot 10^6 \text{ cm/s} .$$

In metals the condition (4) is not satisfied, and hence conditions (1) and (2) cannot be satisfied simultaneously.

Now in semi-conductors the velocity of the electrons may be reduced simply by going to lower temperatures. Upon taking the case of Si, where C is $\sim 1.0 \cdot 10^6$ cm/s, condition (4) becomes

$$T < 160 \text{ °K} .$$

We shall now investigate the relaxation strength to be associated with such relaxation. A free electron gas with a spherical energy surface has a zero shear modulus. Any elastic distortion which leaves the volume unchanged leaves the kinetic energy of the electron unchanged. This would not be true, however, if the quantum numbers of the electrons were to remain unchanged. Thus the change in the kinetic energy per unit volume (K.E.) of the electrons

associated with a shear in P space of the type

$$P_x' = P_x + \gamma P_y$$

$$P_y' = P_y$$

$$P_z' = P_z$$

is

$$\Delta \text{ K.E.} = \tfrac{1}{2}(\tfrac{2}{3} \text{ K.E.})\gamma^2 .$$

If the electrons did not relax to new quantum numbers, their contribution to the shear modulus would thus be $(\frac{2}{3})$ K.E.. The shear relaxation strength would thus be

$$\Delta_G = \frac{(2/3)\,\text{K.E.}}{G} .$$

As an estimate of the value of Δ_G for a typical metal, we shall consider the case of silver. Here the pertinent physical parameters are

$$\text{K.E.} = 3.1 \cdot 10^{11} \text{ erg/cm}^3 ,$$

$$C_{44} = 4.4 \cdot 10^{11} \text{ erg/cm}^3 ,$$

$$\tfrac{1}{2}(C_{11} - C_{12}) = 1.5 \cdot 10^{11} \text{ erg/cm}^3 .$$

The unrelaxed shear modulus $\frac{1}{2}(C_{11} - C_{12})$ would therefore be more than twice as great as the relaxed value. We now pass from a metal to a semiconductor at, say, 100 °K. In order to estimate the relaxation strength under such conditions, we note that the concentration of charge carriers must be below the value at which the electron gas begins to become degenerate. We thereby obtain

$$n \sim T^{\frac{3}{2}}$$

and hence

$$\Delta \text{ K.E.} \sim nT \sim T^{\frac{5}{2}} .$$

Now the degeneracy temperature for silver is 60000 °K. If we therefore reduced the temperature to 100 °K, and also reduced the concentration of electrons so the electron gas would no longer be degenerate, Δ_G in silver would be reduced by a factor of 10^7. Such relaxation strengths are too small to be measured by present techniques. We therefore conclude that it would not be profitable to search experimentally for relaxation associated with the distortion of single Fermi surfaces.

We now consider the case where the Fermi surface is not singly connected, as in the Fig. 13 below. Such a case is of particular importance to us since now τ_m and τ_R need no longer be comparable. In fact it is possible that

$$\tau_R \ll \tau_m .$$

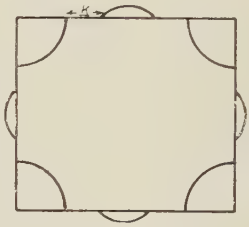


Fig. 13. – A non-singly connected Fermi surface.

Thus let $\boldsymbol{K}$ be the closest distance in K space between the two surfaces. Then in order for an electron interchange to take place via a phonon interaction, the interacting phonon must have a frequency ν of at least $s \cdot \boldsymbol{K}$, where s is the velocity of sound. At low temperatures this interaction will become very small, in fact

$$\tau_R \sim \exp\left[\frac{h\nu}{kT}\right] .$$

The second type of electron relaxation we shall consider is that associated with changes in the equilibrium number of holes and electrons induced by strain dilation.

The τ associated with such a relaxation is essentially the lifetime of minority carriers, and ranges from micro to milliseconds, depending on the purity of the specimens. There should therefore be no difficulty, through proper preparation of the specimen, in passing from the essentially unrelaxed vibrations at $\omega\tau \gg 1$ through $\omega\tau \cong 1$ to essentially relaxed vibrations at $\omega\tau \ll 1$.

We shall next investigate the order of magnitude of the strength of such relaxations. Towards this end we start with a unit volume having the equilibrium number of holes and electrons. We next include the dilation θ by means of an externally applied stress, the number of electrons and holes being kept constant. We finally allow the number of holes and electrons to relax to their new equilibrium values, and estimate the change in dilation θ associated with this relaxation. The final quantity for which we wish an estimate is the ratio

$$\delta\theta/\theta = \Delta_K .$$

If we denote by δn the number of electrons per unit volume transferred from the valence to the conduction band during the relaxation, and δP the difference between the internal pressure exerted by one electron in the conduction and valence bands, then

$$\delta\theta = \delta n \, \delta P / K ,$$

where K is the bulk modulus. Now an estimate of δP is

$$\delta P = \partial E_G/\partial\theta\,,$$

while

$$\delta n = (\partial n/\partial E_G)(\partial E_G/\partial\theta)\,.$$

Hence

$$\Delta_K = (\partial n/\partial E_G)(\partial E_G/\partial\theta)^2/K\,.$$

For an intrinsic semiconductor

$$n = 2(2/\lambda_T)^3 \exp[-E_G/2kT]\,,$$

where λ_T is the wave length given by

$$\lambda_T = h/\sqrt{mkT}\,.$$

We thus obtain

$$\Delta_K = \left\{\frac{4kT}{K(\lambda_T/2)^3}\right\}\left\{\left(\frac{E_G}{2kT}\right)^2 \exp\left[-\frac{E_G}{2kT}\right]\right\}\left\{\frac{\partial \ln E_G}{\partial\theta}\right\}^2\,.$$

The last factor is of the order of magnitude of unity. The middle factor has a maximum value which is also of the order of magnitude of unity, namely $4\exp[-2]$, or 0.54. We are therefore primarily concerned with the first factor.

Without further calculation we can conclude that the first factor is extremely small. Thus kT is very small compared to $K\Omega$, where Ω is an atomic volume, and λ_T^3 is many orders of magnitude greater than Ω. Specifically, at room temperature

$$kT = 4.11\cdot10^{-14}\ \text{erg}\,,$$

$$K = 1.0\cdot10^{12}\ \text{erg/cm}^3\,,$$

$$\lambda_T = 107\cdot10^{-8}\ \text{cm}\,,$$

giving for the first factor the value of $1.3\cdot10^{-6}$. We thus conclude that the relaxation in semiconductors associated with dilation induced changes in electron and hole concentrations has a strength too small to be measured by present techniques.

REFERENCES

[1] This paragraph follows closely the following two papers by the author: *Phys. Rev.*, **52**, 230 (1937); **53**, 90 (1938).
[2] B. S. Berry: *Journ. Appl. Phys.*, **26**, 1221 (1955).
[3] R. H. Randall, F. C. Rose and C. Zener: *Phys. Rev.*, **56**, 343 (1939); K. M. Entwistle: *Journ. Inst. Metals*, **75**, 97 (1948).
[4] The following three paragraph follow closely a treatment given by the author in *Imperfections in Nearly Perfect Crystals* (New York, 1952), Chapter 11.
[5] C. Wert: *Phys. Rev.*, **79**, 601 (1950).
[6] C. Zener: *Phys. Rev.*, **60**, 906 (1941).
[7] T. S. Kê: *Phys. Rev.*, **71**, 533 (1947).
[8] T. S. Kê: *Trans. AIME*, **176**, 448 (1948).
[9] T. S. Kê: *Journ. Appl. Phys.*, **19**, 285 (1948).
[10] T. S. Kê: *Trans. AIME*, **188**, 575 (1950); T. S. Kê and C. Zener: *Chinese Journal of Physics*, **8**, 131 (1951).
[11] J. Friedel, C. Boulanger and C. Crussard: *Acta Metall.* **3**, 380 (1955).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Dislocations in Ionic Crystals.

S. AMELINCKX

University of Ghent - Ghent

1. – Introduction.

In this paper we will discuss the geometry of dislocations in ionic crystals and we intend to demonstrate how observation has confirmed many details predicted by theory. A short review of recent work will be presented, although no attempt is made to give a complete account.

The simple ionic crystals are well suited not only for the study of dislocations in these compounds, but also to be used for a « Model Versuch » concerning dislocations in metals. In the alkali halides for instance the configuration of Burgers vectors is the same as in the face centered metals, and their dislocation geometry should as a consequence not be too different. The major reason for their choice as a material lies in their obvious advantages as compared to metals: they are transparent to visible light so that one can look into them, and it is easy to prepare uncontaminated surfaces by cleavage, which allow the easy application of etching techniques.

The silver halides have been the first to be studied making use of their unique photosensitive properties [1]. The alkali halides were, however, most intensely studied [2-6] because they are easily available as monocrystals. Recently calcium fluoride has proved also to be a suitable material for this kind of study [7-9].

2. – Dislocations in the simple ionic structure.

2·1. *The* NaCl *structure.* – Macroscopic deformation tests lead to the glide elements (110), $[1\bar{1}0]$; which is in accordance with the fact that stable dislocations have Burgers vectors which connect neighbouring ions of the same sign $(a/2)[110]$.

There is some reference in the older literature to other glide planes, (100) and (111), which would operate in special circumstances. Recently GILMAN and JOHNSTON [6] have found in special cases limited dislocation movement in (100) planes in LiF. The geometry of glide elements in the NaCl structure has a striking peculiarity; it is practically impossible to operate by deformation less than two glide systems simultaneously.

For the silver halides, pencil glide is very often found. In NaCl there is evidence for prismatic glide in the [110] direction, at high temperature.

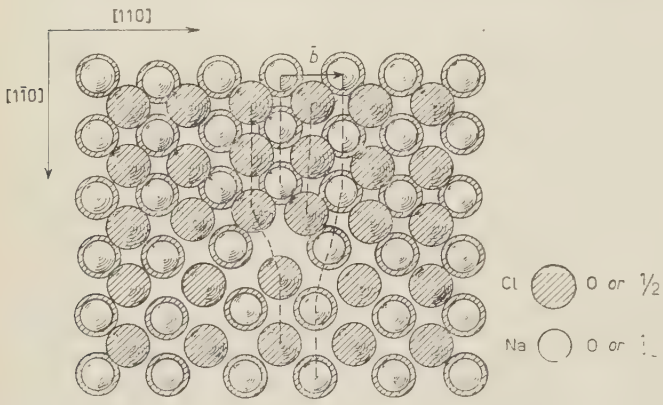


Fig. 1. – Edge dislocation with (110) glide plane and $(a/2)[110]$ Burgers vector in the NaCl structure.

There is some indication (see Sect. 6·1.2) that $a[100]$ dislocations are also stable and can be generated by the reaction $(a/2)[110] + (a/2)[1\bar{1}0] \rightarrow a[100]$.

We present here a schematic view (Fig. 1) of the pure edge dislocation with $(a/2)[110]$ Burgers vector and (110) slip plane and of the pure screw dislocation with the same Burgers vector in the sodium chloride structure (Fig. 2). For the edge dislocation we represent only one plane; the ions in the second plane have exactly the same positions, but opposite sign; the third plane is again the same as the first, etc. It is clear that the edge dislocation can be thought of as made by introducing *two* supplementary half planes of the (110) type, whilst the screw dislocation consists of *two* interleaved helicoidal surfaces. The two half planes are necessary to maintain charge balance. For the same reason the half planes are adjacent; the dislocation is not extended like in the f.c.c. metals; the energy of a twin interface being probably much too high.

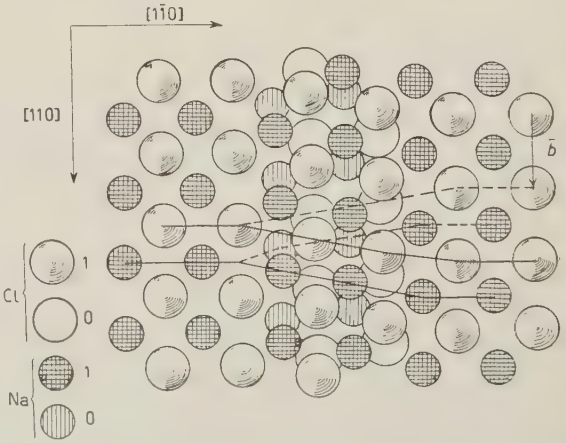


Fig. 2. – Screw dislocation in the NaCl structure.

Detailed calculations, to determine the configuration of ions in the dislocation core and the energy of dislocations in the NaCl structure, have been made by HUNTINGTON, DICKEY and THOMSON [10]. They find the results summarized in the following Table I.

TABLE I. – *Elastic energies at* 80 °K *and* 1000 °K *of dislocations in* $(1\bar{1}0)$ *slip plane as a function of the angle* θ *between the normal to the line and the Burgers vector* $(a/2)$ [110] ($\theta=0$: *edge*; $\theta=90°$: *screw*).

$E_{80} = (0.454 + 0.061 \cos 2\theta) \ln (R/r_0)$
$E_{1000} = (0.214 + 0.032 \cos 2\theta) \ln (R/r_0)$
R: Outer cut-off radius (or the distance between dislocations).
r_0: Inner cut-off radius (size of the core of the dislocation).
Units are eV/atomic distance.

It turns out that the screw dislocation has the smallest energy.

The same authors have also tried to derive on theoretical basis whether the a[100] dislocation would be stable. They find, taking into account the elastic energy only, that a a[100] edge dislocation will *not* be stable with respect to dissociation into two mixed (45°) a[110] dislocations. On the other hand the a[100] screw dislocation would be stable. This seems to be confirmed by our observations (see further Sect. 6·1.2).

2·2. *The* CaF_2 *structure.* – Much less work has been done about the dislocations in this kind of structure. The stable dislocations have Burgers vectors $(a/2)$[110], and in the same circumstances as in the NaCl structure the a [100] will presumably be stable also. The glide plane reported in the literature [11] is (001). A schematic view of the edge dislocation with a (001) glide plane and Burgers vector $(a/2)$[110] is shown in Fig. 3; it contains again two supplementary half planes.

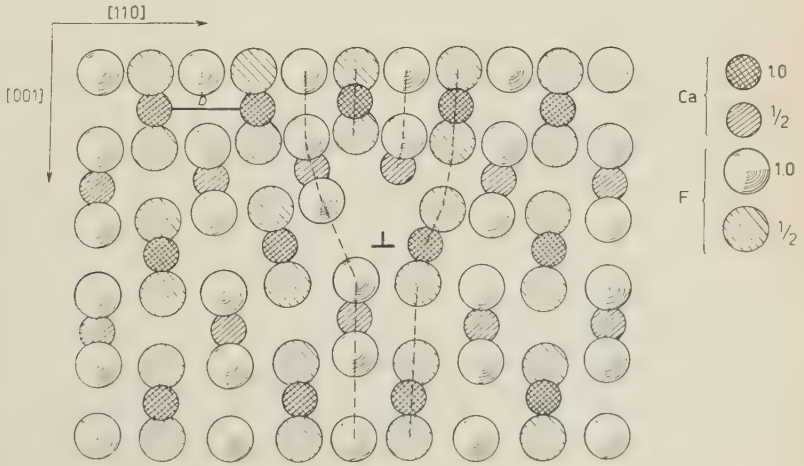


Fig. 3. – Edge dislocation in the CaF_2 structure. The glide plane is (001) and the Burgers vector is $(a/2)$ [110].

2 3. *The* CsCl *structure.* – In this structure the configuration of Burgers vectors is the same as in the primitive lattice; the smallest Burgers vectors are $a[100]$ and connect again the nearest ions of the same kind. The stability of $a[110]$ and $a[111]$ is doubtful although it is to be expected for reasons of analogy with NaCl that $a[110]$ should occasionally form. The glide planes

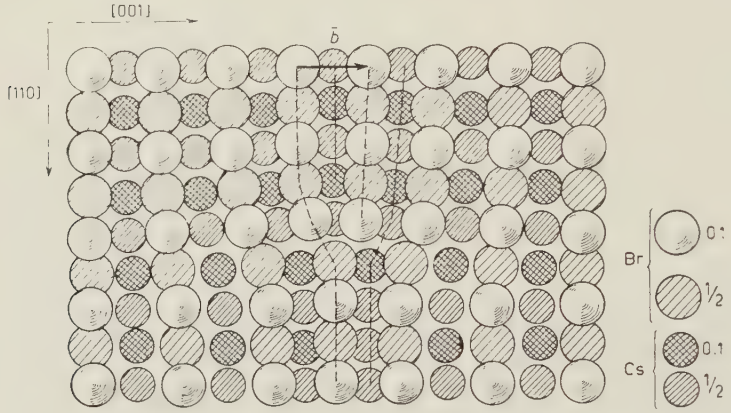


Fig. 4. – Edge dislocation in the CsBr structure. The Burgers vector is $a[100]$, the glide plane (110).

in this structure appear to be (011). A schematic representation of the edge dislocation in the CsCl structure, having a (100) as a Burgers vector and (011) as a glide plane is given in Fig. 4. Pencil glide is possible in this structure, as two different glide planes pass throught the glide vector.

3. – Specific properties of dislocations in ionic crystals.

3·1. *Effective charges.* – In ionic crystals charges are associated with jogs and other discontinuities in dislocations. Seitz and Mott [12] have independently shown that an effective charge of $e/2$ is associated with a jog. This can be demonstrated quite simply by considering a situation as shown in Fig. 5 which refers to the edge of a dislocation in the NaCl structure. It is clear that there is one negative charge in excess; from symmetry it is

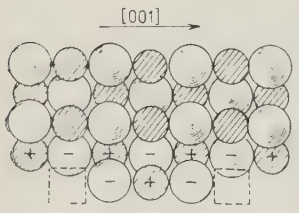


Fig. 5. – Edge of supplementary half planes of the edge dislocation of Fig. 1. Note the incipient vacancies surrounded by squares; they have an effective negative charge.

clear that this is equally distributed between the two endpoints of the row, i.e. between the two jogs. It is immediately clear that the addition of one positive ion to the row displaces the jog sideways and inverts the sign of the effective charge at the jog. This again shows that the effective charge is $\pm e/2$. Similar jogs can be considered in the other ionic structures; Fig. 6 shows the situation in the CsCl structure (*). According to SEITZ [12] this effective charge is responsible for speck formation during the print out process in the Ag-halides.

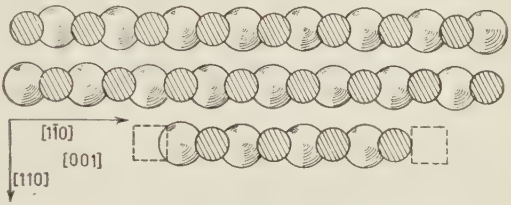


Fig. 6. – Edge of the supplementary half planes of the edge dislocation of Fig. 4.

The same concept was also used to explain the coagulation of F-centers into colloids along dislocations in additively coloured NaCl crystals [2, 3]. SEEGER [13] has pointed out that the jogs, referred to above, are different from the ones formed by the intersection of screw dislocations; the latter kind of jog contains two rows of ions and has *no* effective charge. An activation energy of the order of 0.2 eV is necessary to dissociate this jog into two charged jogs.

Jogs are the preferred sites where point defects are created or annihilated, SEITZ calls them « incipient vacancies ». Incipient vacancies are shown in Fig. 5 by means of the squares.

Effective charges can be associated to other discontinuities in dislocation lines. Experimental evidence presented by KANZAKI [14] suggests that the

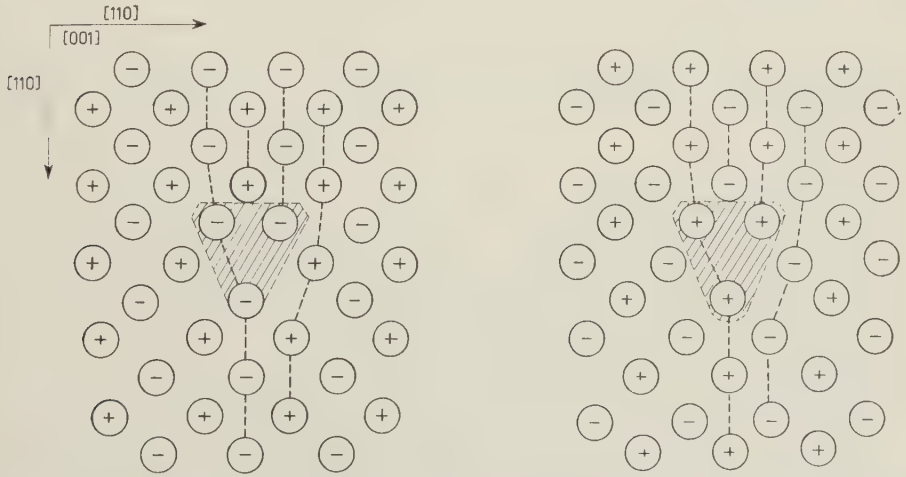


Fig. 7. – To illustrate the effective charges associated with the emergence points in the (001) plane of edge dislocations in the sodium chloride structure.

(*) It is clear that again an effective charge is associated with the jogs.

emergence points of edge dislocations in the (100) plane but *not* in the (110) plane of the NaCl structure are effectively charged. This can be demonstrated intuitively by means of Fig. 7 and Fig. 8. The first shows the two kinds of emergence points in a (001) plane, one positively and one negatively charged. The conservative motion of the dislocation does not change the effective charge at the emergence point. On the other hand the addition of one row of ions, by climb e.g., inverts the sign of the effective charge, which suggests that the site above the emergence point has an effective charge of about $e/2$. On the other hand, the emergence point in (110) (see Fig. 8) does not present any singularity for the charges.

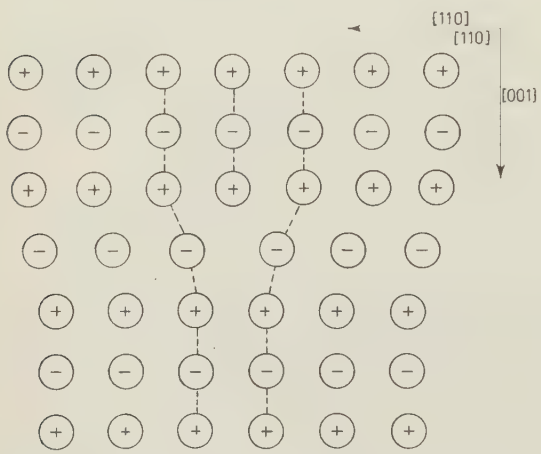


Fig. 8. – To illustrate the emergence point in the (110) plane of an edge dislocation in the sodium chloride structure.

According to KANZAKI also node points of the dislocation network should have an effective charge.

3·2. *Dislocations and ionic conductivity.* – GUYLAI and HARTLY [16] have found a sudden increase, by a factor of the order of 100, in the conductivity of NaCl crystals after deformation at 39°. The increased conductivity decays in about 20 min at room temperature. New pulses can be produced by further plastic flow. The explanation advanced by SEITZ [17] considers the rise in conductivity as being due to an increase in the number of carriers for ionic conductivity. These carriers, vacancies of both signs, would be produced by the intersection of dislocations. Although the vacancies are initially produced in pairs, they would disperse as a consequence of the local increase in temperature as the dislocation passes. The positive ion vacancies should have sufficient mobility at room temperature to drift afterwards to negative ion vacancies and form neutral pairs, which no longer contribute to the conduction process. This would explain the rather slow decay.

JOHNSTON [18] has found in AgBr a permanent increase in ionic conductivity, after plastic deformation. The increase varies linearly with strain (for small strains). The anisotropy of the effect indicates that it is greater parallel to the slip plane than perpendicular to it. The activation energy for conduction in the disordered region is less than 4600 cal/mol compared to 18450 cal/mole for the intrinsic conductivity. The regions of disorder are assimilated with piled up groups of dislocations.

4. – Dislocation nets in the simple ionic structure.

We will now derive what one has to expect for the geometry of dislocation nets in these structures [19, 20]. Experiment has shown that one and two dislocation boundaries form the major part of the networks. Dislocations of a third kind can generally be treated as singularities in two dislocation nets.

4·1. *One dislocation boundaries.* – If only one kind of dislocations is present in a boundary, Frank's formula [21] shows that the plane of the boundary is perpendicular to the Burgers vector. If the dislocations are confined to their glide planes the rotation axis is uniquely determined also; it is the line of intersection of the glide plane and the boundary plane. If climb can occur the rotation axis can have any orientation, provided, it is in the boundary plane. Table II gives the characteristics of one dislocation boundaries in the simple ionic structures.

TABLE II.

Crystal	Burgers vector	Glide plane	Boundary plane	Rotation axis = direction of lines	
				climb	glide
NaCl	$(a/2)[1\bar{1}0]$	(110)	$(1\bar{1}0)$	$[uuw]$	[001]
CaF_2	$(a/2)[1\bar{1}0]$	(001)	$(1\bar{1}0)$	$[uuw]$	[110]
CsCl	$a[100]$	(011)	(100)	$[0vw]$	$[01\bar{1}]$

4·2. *Two dislocation boundaries.* – We call the Burgers vectors of the two sets of dislocations $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, and the unit vectors perpendicular to the glide planes $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, $\boldsymbol{\omega}$ is the rotation vector and $\boldsymbol{\nu}$ the unit normal on the contact plane. One has to distinguish two cases.

4·2.1. The dislocation lines are all parallel. – Using Frank's formula one can show that $\boldsymbol{\omega}$ is parallel to the direction of the lines and that $\boldsymbol{\nu}$ is parallel to $\boldsymbol{\omega} \times (\boldsymbol{b}_1 \times \boldsymbol{b}_2)$, i.e. the boundary plane is the plane determined by $\boldsymbol{\omega}$ and $\boldsymbol{b}_1 \times \boldsymbol{b}_2$. If dislocations have to be confined to their glide planes, $\boldsymbol{\omega}$ has the direction of $\boldsymbol{a}_1 \times \boldsymbol{a}_2$ and the direction of the unit normal on the boundary plane is then given by $(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \times (\boldsymbol{b}_1 \times \boldsymbol{b}_2)$.

The boundaries are *pure tilt* boundaries. For the particular structures considered here one obtains as possible boundaries the tilt boundaries in Table III.

4·2.2. The dislocation lines are not parallel. – It can be shown that the rotation vector $\boldsymbol{\omega}$ is unambiguously determined: it is parallel to $\boldsymbol{b}_1 \times \boldsymbol{b}_2$; the boundary plane has a certain degree of freedom, it is only subject

TABLE III. – *Two Dislocation Boundaries.*

Glide elements of dislocations	Climb allowed				Pure glide			
	ω	**ν**	Type	Direction of lines	**ω**	**ν**	Type	Direction of lines
NaCl (*)								
$\bar{a}_1=[1\bar{1}0]$; $\bar{b}_1=[110]$	$[uvw]$	$[hk0]$	tilt $(hu+kv=0)$	$[uvw]$	[001]	$[kh0]$	tilt	[001]
$\bar{a}_2=[110]$; $\bar{b}_2=[1\bar{1}0]$	[001]	$[hkl]$ $(l\neq 0)$	mixed	$[l,\ \bar{l},\ k-h]$ $[\bar{l},\ \bar{l},\ h+k]$	[001]	$[khl]$ $(l\neq 0)$	mixed	$[l,\ \bar{l},\ k-h]$ $[l,\ l,\ h+k]$
Special case	[001]	[001]	pure twist	[110], $[1\bar{1}0]$	[001]	[001]	pure twist	[110], $[1\bar{1}0]$
$\bar{a}_1=[110]$; $\bar{b}_1=[1\bar{1}0]$	$[uvw]$	$[hkl]$	tilt	$[uvw]$	$[1\bar{1}\bar{1}]$	$[01\bar{1}]$	tilt	$[1\bar{1}\bar{1}]$
$\bar{a}_2=[101]$; $\bar{b}_2=[10\bar{1}]$		$h+k+l=0$	$hu+kv+lw=0$					
	[111]	$[hkl]$ $h+k+l\neq 0$	mixed	$[\bar{l},\ \bar{l},\ h+k]$ $[k,-(h+l),k]$	[111]	[011]	mixed	$[\bar{1}1\bar{1}]$, $[11\bar{1}]$
Special case	[111]	[111]	pure twist	$[\bar{1}\bar{1}2]$, $[1\bar{2}1]$	—	—	—	—
CaF_2 (*)								
$\bar{a}_1=[001]$; $\bar{b}_1=[\bar{1}10]$	$[uvw]$	$[hk0]$	tilt $(hu+kv=0)$	$[uvw]$	$[\bar{k}h0]$	$[hk0]$	tilt	$[\bar{k}h0]$
$\bar{a}_2=[001]$; $\bar{b}_2=[1\bar{1}0]$	[001]	$[hkl]$ $(l\neq 0)$	mixed	$[l,\ \bar{l},\ k-h]$ $[\bar{l},\ \bar{l},\ h+k]$	[001]	[001]	pure twist	[110], $[1\bar{1}0]$

$\bar{a}_1=[001]$; $\bar{b}_1=[1\bar{1}0]$ $\bar{a}_2=[010]$; $\bar{b}_2=[10\bar{1}]$	$[uvw]$	$[hkl]$ $h+k+l=0$	tilt $hu+kv+lw=0$	$[uvw]$	[100]	[011]	tilt	[100]
	[111]	$[hkl]$ $h+k+l\neq 0$	mixed	$[\bar{l},\ \bar{l},\ h+k]$ $[k,\ -(h+l),\ k]$	[111]	[100]	mixed	[001] [010]
Special case	[111]	[111]	pure twist	$[\bar{1}\bar{1}2]$, $[1\bar{2}1]$				
CsCl (+)								
$\bar{a}_1=[01\bar{1}]$; $\bar{b}_1=[100]$ $\bar{a}_2=[101]$; $\bar{b}_2=[010]$	$[uvw]$	$[hk0]$	tilt $hu+kv=0$	$[uvw]$	$[11\bar{1}]$	$[1\bar{1}0]$	tilt	$[11\bar{1}]$
	[001]	$[khl]$ $(l\neq 0)$	mixed	$[0\bar{l}k]$, $[l0\bar{h}]$	[001]	[001]	twist	[100], [010]
Special case	[001]	[001]	pure twist	[100], [010]	—	—	—	—
$\bar{a}_1=[011]$; $\bar{b}_1=[100]$ $a_2=[011]$; $b_2=[100]$ (Pencil glide)	$[0cw]$	[100]	tilt (Boundary is effectively one dislocation boundary)	$[0cw]$	$[0cw]$	[100]	tilt	[011], [011]

(*) Burgers vectors are in units $a/2$.
(+) Burgers vectors are in units a.

to the condition $\nu \cdot (b_1 \times b_2) \neq 0$. The dislocation lines with Burgers vector b_1 and b_2 have respectively the direction $b_2 \times \nu$ and $b_1 \times \nu$. In the planes for which $\nu \cdot (b_1 \times b_2) = 0$ the boundaries are pure twist boundaries; in the general case the boundaries have mixed character.

If dislocations are confined to their glide planes ν has to be parallel to $(a_1 \times b_2) \times (a_2 \times b_1)$.

Table III summarizes the characteristics of the boundaries of this type for the specific structures here considered.

4·3. *Stability considerations.* – The simple geometrical models just derived do not necessarily represent stable configurations. We will now discuss the actual shape, towards which the described configurations will evolve.

Consider first the nets with a [111] rotation axis in the NaCl structure. The geometrical model in the simplest case of a (111) boundary plane is a lozenge shaped net as shown in Fig. 9(*a*) in dotted line; this is however not stable. The elastic energy associated with the net can be decreased by the formation of additional segments of dislocation according to the reaction $(a/2)[1\bar{1}0] + (a/2)[\bar{1}01] \rightarrow (a/2)[011]$, in the way shown in Fig. 9(*a*) by means of full lines. This process not only reduces the total length of dislocation line by a factor $\sqrt{3}/2$ but also brings all the dislocations involved into the pure screw orientation, which has the smallest energy per unit length [10]. If the contact plane is not (111) it is easy to check that the net is such that its orthogonal projection on (111) is again the just described net. For the special case of a (100) plane, which is the usual observation plane in the case of the NaCl structure, the net has the shape shown in Fig. 9(*b*) according to the projection rule. The short segment remains a pure screw dislocation, whilst the other segments acquire mixed character, the ratio c/a being 3. Minimizing the total line energy of the net leads to a value for the angles at node points which is in accordance with this shape.

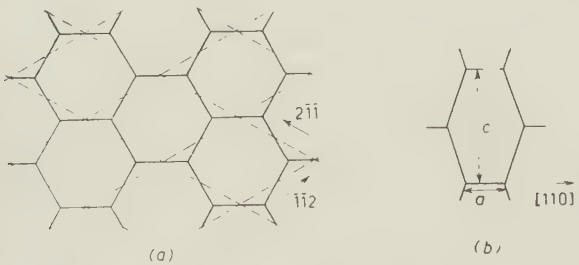


Fig. 9. – *a*) Formation of the hexagonal network in a (111) plane. *b*) Shape of meshes of the hexagonal network in the cube plane.

The net with $a[001]$ rotation axis and $a(100)$ contact plane is a square net with lines parallel to [110] and $[1\bar{1}0]$; it is stable to a first approximation. There is however a slight tendency for the formation of additional segments of a [100] dislocation at the intersection points by means of the reaction $(a/2)[110] + (a/2)[\bar{1}10] \rightarrow a[100]$.

This reaction may diminish the elastic energy for certain orientations of the lines involved. This tendency becomes more pronounced if the plane of the net is inclined with respect to (001), because the meshes then become lozenges and the two segments which have tendency to combine, come nearer to the parallel orientation; they acquire moreover an orientation of higher energy. These nets may as a consequence also become hexagonal. The difference with the previously described nets should however be evident from the difference in angles at the node points. Whereas in the first case angles are all approximately equal to 120°, in the present case one angle should be much smaller than the two others (see photograph).

For the CaF_2 structure similar considerations apply; observation of nets in (111) planes is now favoured as a consequence of the (111) cleavage. Hexagonal nets should now be observed undeformed whereas the nets with a $a[001]$ rotation axis and $a(111)$ boundary plane would have rectangular meshes, their lines being parallel to $[1\bar{1}0]$ and $[11\bar{2}]$ and the ratio of the sides being $\sqrt{3}$. Small segments of $a[100]$ dislocation would probably form at the intersection points.

In the CsCl structure one would expect square nets in the cube plane, the lines being pure screws parallel to the cube edges. The same argument used above allows to predict that in oblique planes the meshes may become hexagones by the formation of segments $a[110]$ according to the reaction $a[100] + a[010] \rightarrow a[110]$. The angle at the node points should again deviate appreciably from 120°. The observation of threefold nodes in a CsCl type crystal would be sufficient to conclude to the stability of the $a[110]$ dislocations.

4·4. *Junctions of grain boundaries.* – The cellular substructure of single crystals is built up by nets of the type just discussed. The main features in this structure are nodal lines, where three boundaries intersect, and intersection points of four (usually non-coplanar) nodal lines. Under certain simplifying assumptions the geometry of grain boundary junctions has been studied in the NaCl structure [20]. It turns out that there is only a limited number of ways of joining three dislocation nets in a way which satisfies the following requirements:

(i) The three boundary planes have to intersect along a line.

(ii) Rotations have to be conserved.

(iii) The dislocation content has to be conserved, i.e. no dislocation lines can end at the junction line.

(iv) Every boundary has to have a stable configuration so that no far reaching strains remain.

Relations have also been derived between the densities of emergence points, in a given plane, of the dislocations in three intersecting grain boundaries. For example for three tilt boundaries with Burgers vectors [110] and [1$\bar{1}$0] the following relation should exist between the densities of emergence points ϱ_i in the (001) plane

$$\sum_{i=1}^{3} \varrho_i/(\cos\varphi_i + \sin\varphi_i) = 0\,, \tag{4·4.1}$$

where φ_i is the angle between the trace of the boundary plane and [110]. Relations of this type can easily be checked by means of etch pit counts [22-24].

5. – Observation methods.

We will only focus attention on such methods which can be termed directly.

5·1. *Surface methods.* – In principle growth and evaporation phenomena can be used to locate emergence points of dislocations, as centers of growth or evaporation spirals. Although they have effectively been used for the alkali halides [25, 26] it should be noticed that they can only be applied to dislocation with a screw component. They furthermore necessitate the use of very well prepared growth or evaporation surfaces and refined optical methods.

The study of the geometry of cleavage faces can also be of interest and provides [25, 27] e.g. a simple mean of distinguishing tilt and twist boundaries: to twist boundaries cleavage steps are necessarily attached, whilst this is not the case for tilt boundaries.

The most convenient surface method however consists in producing etch pits at the emergence points. Ionic crystals have special advantages in that it is easy to produce an uncontaminated surface by cleavage. In these circumstances a suitable etchant is able to attack immediately the stressed region around the emergence point of the dislocation and produce an etch pit, without necessitating the presence of some concentration of impurities around it. Such etchants will then preferentially etch freshly formed dislocations. Dislocations with an atmosphere will be attacked less because the stresses have been relieved partially. Other etchants may attack the impurity region around grown in dislocations and they will then less well mark newly formed dislocations.

Etch pits have been used since many years to study the symmetry of crystals and useful information concerning the composition of etchants can be found in the older literature.

They have, however, only these last years been explicitely used to detect

emergence points of dislocations. Some etchants suitable to reveal dislocations in specific ionic crystals are tabulated here:

TABLE IV.

Substance	Etchant	Etching time	References
NaCl	Anhydrous Methylalcohol	As short as possible (attacks better « grown in » dislocations)	AMELINCKX [22]
AgCl } AgBr }	3.5 n aqueous solution of Hyposulphite		MITCHELL [28]
LiF	1 part HF conc. 1 part glacial acetic acid + 1 vol. per cent of conc. HF saturated with FeF_3	30-60 seconds (attacks all disl.)	GILMAN and JOHNSON [6]
	aqueous solution of FeF_3 ($1.5 \cdot 10^{-4}$ molar)	2 min. (attacks better fresh dislocations)	»
$CaCO_3$	diluted acetic acid	30 seconds	—
CaF_2	concentrated H_2SO_4	10-30 minutes	BONTINCK [9]

In contrast with the decoration methods to be described in the next section, the etch pits do not pin the dislocation. It is thus possible to observe the displacement of a dislocation by successive etching.

An easy check on the effectiveness and reproducibility of an etch is obtained by comparing the etch pattern of two halves of a cleaved crystal: the patterns should be mirror images apart from small detail. This was found to be the case for NaCl [2, 22] and LiF [6]. This allows to have two identical surfaces for what concerns their dislocation configuration; one of them can be subject to some treatment and compared with its « replica ».

Continued etching, eventually with intermediate polishing, allows to follow the dislocation line in depth.

The one to one correspondence between decorated dislocations and the etch pits centered on them, has been verified for NaCl [2] and AgBr [28].

5·2. *Decoration methods.* – Direct observation of the dislocation lines, using decoration methods is best suited for a detailed verification of all geometrical aspects of the theory. This method has been applied with success to AgBr,

TABLE V.

Crystal	Method of decoration	Precipitating entity	Nature of particles	Observation method	References
1. AgBr pure AgCl pure	Printing out by exposure to light in the tail of the absorption band	Interstitial silver	Silver	Dark field	[1, 29, 14]
2. NaCl pure	Strong additive coloration with sodium metal	*F* centers or aggregates	Sodium	Ultramicroscopy	[2, 4]
3. NaCl+ 0.01% AgCl	Additive coloration with sodium metal		Silver	»	[3]
4. NaCl+ 3% AgCl	Anneal in hydrogen		Silver	»	(31, 30]
KCl+ 0.75% AgCl	»	»	»	»	
KBr+ 1% AgBr	»		»	»	
5. NaCl pure KBr pure	Diffusing in of Au by decomposition of the corresponding gold halide in vacuum	Interstitial gold	Gold	Transmission	[5]
6. CaF_2	Additive coloration with Ca		Calcium	Ultramicroscopy	[7, 8]
	Annealing in air		?	»	

AgCl [1, 14, 29], NaCl [2-5, 31], KCl and KBr [5, 30]. All decoration methods are based on the segregation along the dislocation line of some entity which forms observable particles. This implies the introduction of, either some impurity, or an excess or defect of one of the constituents of the crystal. Table V summarizes the characteristics of some decoration methods which have been successful in revealing dislocations. For the details of the methods we refer to the original publications.

These methods also bring evidence for the interaction of point defects and dislocations.

6. – Results of observations.

6·1. *Dislocation nets.* – We will now discuss some typical results obtained by the various methods referred to above.

6·1.1. *Etch patterns.* – Prior to the use of decoration methods, etching has been used to study the geometry of dislocations in NaCl and to mesure their density. Usually values of the order of 10^5 or 10^6 per cm^2 are found for annealed NaCl crystals. Carefully prepared single crystals of NaCl may contain as little as $10^4/cm^2$ dislocations; the same applies to LiF, for which $10^4/cm^2$ dislocations were found [6]. The number of pits within the subcells is an order of magnitude smaller than the number of pits in the boundaries.

A detailed study of the geometry of pits along the boundaries may reveal the character of these boundaries. The particular shape of a pit gives information concerning the inclination with respect to the observation plane of the line on which it is centered. The sense into which the deepest point of the pit is displaced with respect to the center of the periphery is also the sense of inclination of the line. The assumption is that the pit is on every moment centered on the dislocation line. For a tilt boundary all pits have the same shape; for a twist boundary on the other hand successive pits will have different shapes as shown in Fig. 10. Such etch patterns were observed in NaCl [2].

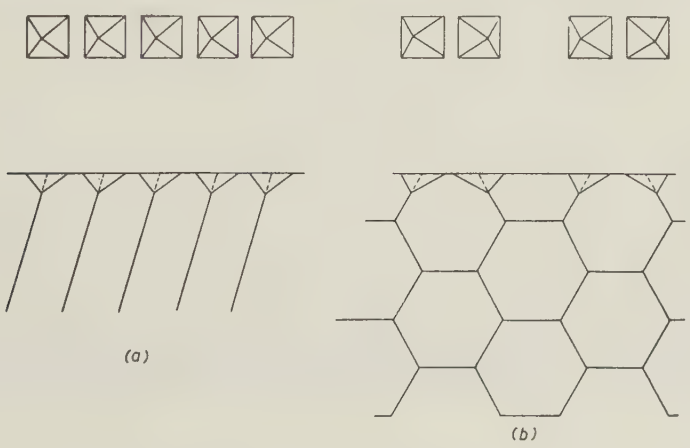


Fig. 10. – Shape of etch pits: *a*) at a tilt boundary; *b*) at a twist boundary.

Measurements of dislocation densities in three meeting boundaries confirming equation (4·4.1) have been made in NaCl [22].

6·1.2. Dislocation nets in NaCl and KCl studied by decoration methods [30]. – The specimens are cleavage lamellae from crystals treated

in the way summarized in Table V, method 4; the plane of observation is the (001) plane. Specimens are observed in ultramicroscopy.

(i) Parallel lines. Sets of parallel lines characteristic of tilt boundaries are observed in a variety of planes and directions, proving that climb is important. Photograph 1 reproduces an example. Specially simple tilt boundaries can be introduced artificially by bending a crystal about an axis parallel to a cube edge and annealing thoroughly. The crystal polygonizes and dislocation walls perpendicular to the glide planes result. The dislocation lines are parallel to the axis of bending. They are the one dislocation boundaries referred to in Table II.

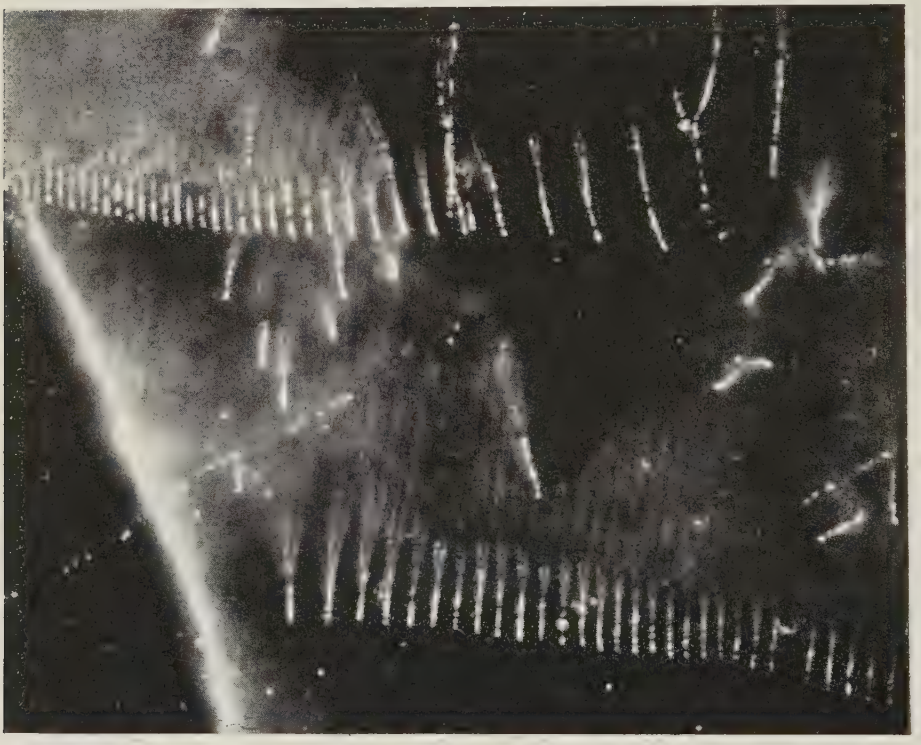


Photo 1. — Tilt boundaries in KCl crystal. The microscope is focussed on their emergence point in the surface.

This was verified in the case of NaCl by means of etching and by decoration methods [2, 4, 5]. Photograph 2 shows a set of polygonization walls. The lines are seen as dots because they are perpendicular to the plane of observation.

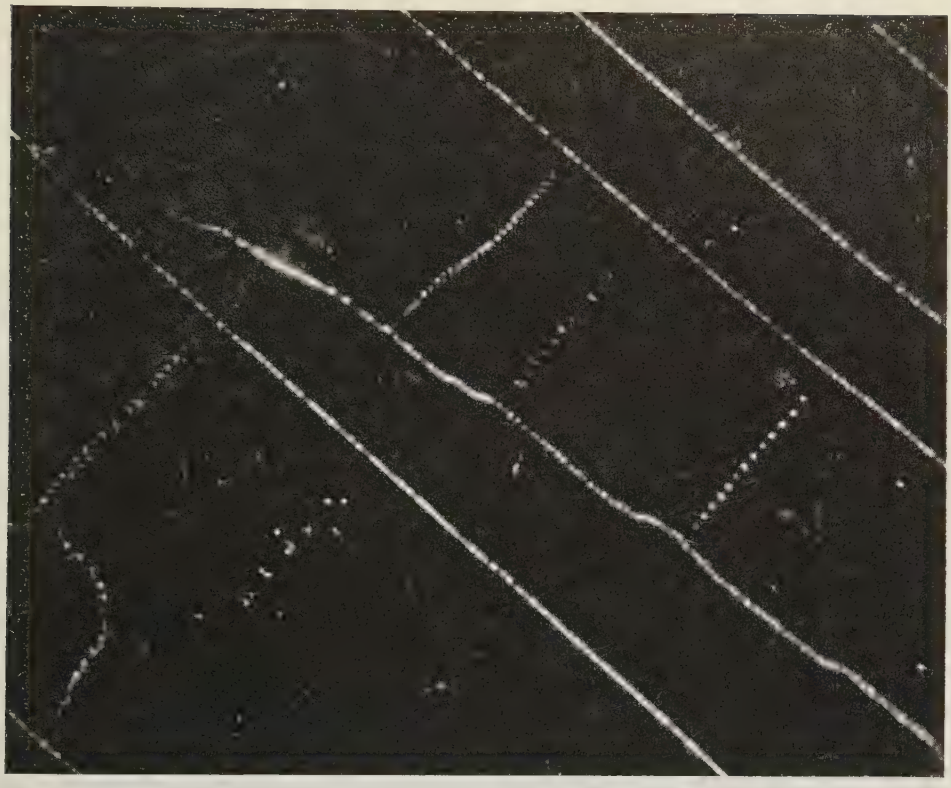


Photo. 2. – Polygonization walls in a bent and annealed NaCl crystal. Axis of bending was perpendicular to the plane of observation.

(ii) Square nets. In accordance with the theoretical predictions square nets were found. When lying in the cube plane the directions of the lines are very approximately [110] and [1$\bar{1}$0]. In general no observable segments of dislocation line with Burgers vector a[100] are formed at the intersection points. Photograph 3 represents an example observed in KCl. Nets of this type lying in oblique planes were also found, and in such cases observable segments of a[100] were often found in the acute angles of the lozenge shaped meshes in accordance with theoretical speculation (photograph 4, see arrows).

We will now consider some typical singularities which are frequently found in square nets. The most common among them are the zig-zag lines as e.g. visible on photograph 5. They originate in dislocations which have a Burgers vector which forms an angle of 120° (or 60°) with the Burgers vector of the set they intersect. Additional segments then form at the intersection points and this transforms the originally straight line into a zig-zag line. The distribution of Burgers vectors is best described by means of a lettering pattern

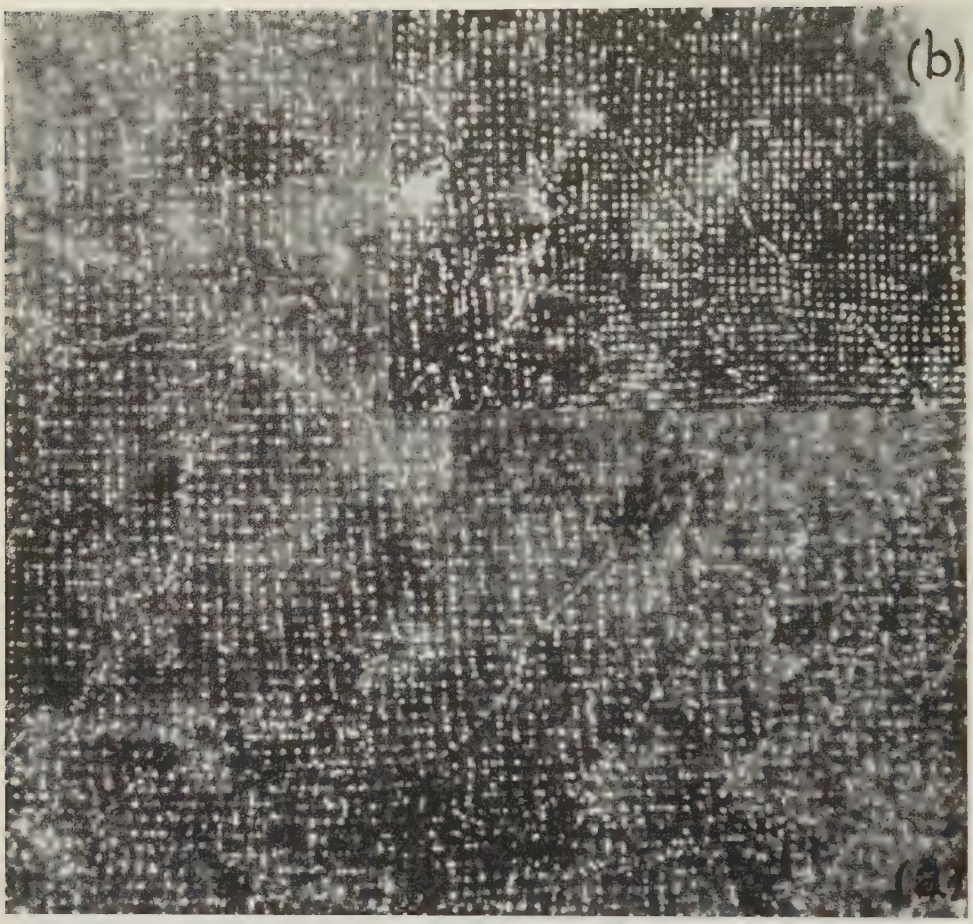


Photo 3. *a*) Square net in KCl crystal. *b*) Square net with preferential decoration at the node points.

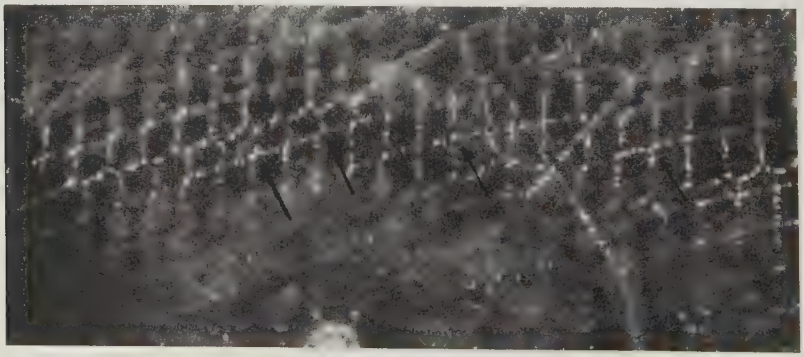


Photo 4. – Square in oblique plane. Note the formation of small segments of a 100 dislocation at the intersection points marked by arrows. Note also the presence of a zig-zag line.

Photo. 5. – Zig-zag lines in square network in KCl. This pattern is analysed in Fig. 11 and 12.

of the kind first used by FRANK [19] and which is based on Thompson's notation for Burgers vectors in the face centered cubic lattice. Such a lettering pattern is given in Fig. 11.

The same photograph also shows another interesting feature. Although no visible segments of $a[100]$ are present at the majority of intersection points of the square net, such segments systematically form at places where the just described zig-zag line jogs from one row of square meshes to the next. We believe that the presence of the singular dislocation line catalyzes the formation of such a segment. To demonstrate how this may happen consider the crossing of the lines CD of the net and of the singular line CA (Fig. 12). Their interaction is such that a torque is exerted which tends to bring the

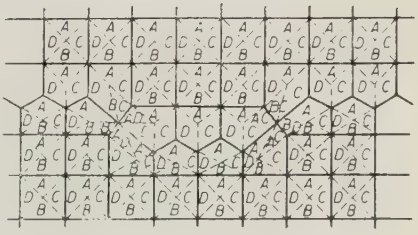


Fig. 11. – Zig-zag lines and the formation of $a[100]$ dislocations in a square network. This figure can be compared with photograph 5.

lines locally into the antiparallel configuration. In the letter notation this means that the dislocations will bend in such a way as to try to form parallel segments in the sector which has the same lettering for both dislocations. This is shown for CD and CA, as well as for BA and CA in Fig. 12. It is now clear that the presence of the singular line CA causes the lines CD and AB to adopt the parallel configuration, which facilitates them to combine. The dislocations CA and BA have adopted hereby an orientation which deviates from the initial screw orientations, and which has a higher energy (see Sect. 2·1); this makes it profitable for the two segments to combine into a segment of approximately pure screw dislocation with $a[100]$ Burgers vector. From the calculation of HUNTINGTON, DICKEY and THOMSON [10] it was to be expected that this dislocation would be stable.

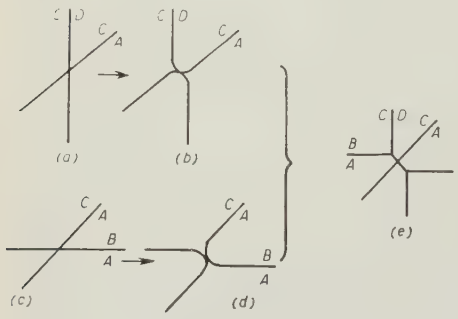


Fig. 12. – To illustrate how the presence of a singular dislocation CA may catalyze the formation of a [100] dislocations at the intersection points of a square network.

On the other hand the line CA is subject to two torques of opposite sense and in the ideal case they would be of the same magnitude, so that finally the line CA would remain straight. As dislocations with Burgers vectors larger than $a[100]$ are not stable in the sodium chloride structure, there is no reason why an additional segment should form at the intersection point of the $a[100]$ dislocation and the CA dislocation. A satisfactory explanation for the observed configuration is thus obtained. A check on its validity is obtained by considering what happens when the zig-zag line jogs back to its original row, as can be observed on photograph 5. From the corresponding lettering patterns of Fig. 12 it is clear that now a line with Burgers vector DA and having a direction roughly perpendicular to the one of the segment CA, crosses the intersection point. A quite similar reasoning as above, shows that there is again no net torque on DA, and that AB and CD are forced into parallel positions in the right directions.

(iii) Hexagonal networks. Hexagonal networks characteristic of $a[111]$ rotation axis were observed in a variety of planes. Of special interest for our purpose are nets lying in a cube plane parallel to the observation plane: such nets can be observed undeformed over a large area. The ratio c/a (see Fig. 9(b)) was measured for a number of nets and the mean value was found to be 2.8 in fair agreement with the theoretical value. Nets of this kind are shown in photographs 6 and 7. It is found that the shortest

segment of the meshes is parallel to the [110] direction, as required by theory. The short segments are, as a consequence, presumably pure screw. It will

Photo 6. – Hexagonal network containing singular lines analysed in Fig. 13 and 14.

be evident from the photograph that along such segments the density of the decoration is systematically smaller. This suggests that screw dislocations are less well decorated.

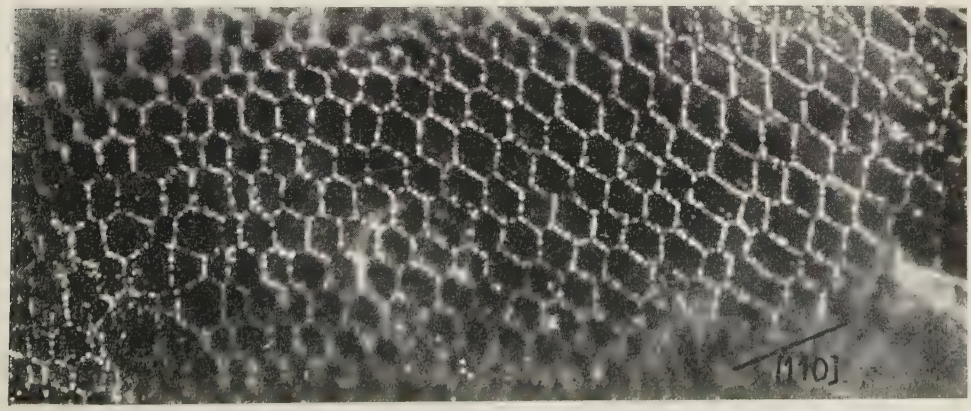


Photo 7. Hexagonal network in KCl.

We will now discuss a few typical singularities which are usually found in hexagonal nets. We first consider the line marked by (1) in photograph 6. It is clear that this line intersects perpendicularly a set of parallel segments of the net. The interpretation is straightforward: the line has a Burgers vector which is perpendicular to the Burgers vector of the intersected set of lines. Such a line may jog from one row of meshes to the next; this happens in the point marked [3]. When analyzing this configuration in terms of Burgers vectors one gets the pattern of Fig. 13(*b*). If this line changes direction and intersects now segments of the net having another Burgers vector, a supplementary row of hexagones results (see 2, photograph 6). In this area nothing would reveal the presence of a singular line if the change in direction were not present. The hypothetical original configuration of dislocation lines Fig. 14(*a*) which would give rise to the observed pattern Fig. 14(*b*) is shown also.

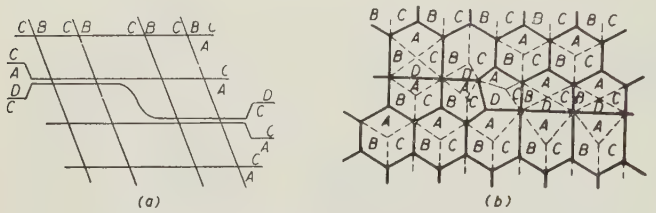


Fig. 13. – *a*) Configuration of lines giving rise to the pattern. *b*) Compare with photographs 6, arrows (1) and (3).

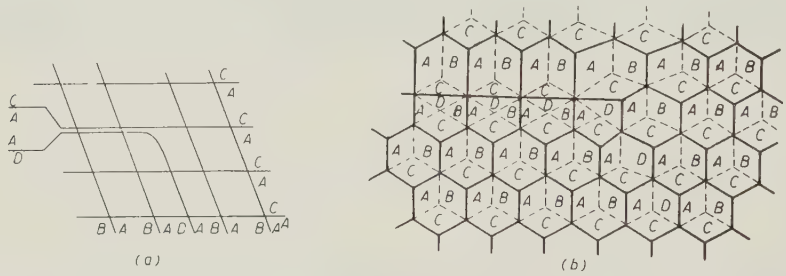


Fig. 14. – *a*) Configuration of lines giving rise to the pattern. *b*) Compare with photograph 6, arrow (2).

(iv) Other singularities in nets. We will now discuss in some detail the pattern shown in photograph 8. This pattern may be considered as a square net containing singular lines. In the left top part the singular line with Burgers vector AC intersects the node points of the square net. As a results segments of $a[100]$ dislocation are formed, according to the mechanism discussed before (see (ii)). Afterwards the same line forms a zig-zag line which bends over 90° downwards and now forms a zig-zag line in the direction of the second set of dislocations of the square net. It apparently

ends at the point where two curvilinear hexagones are visible. A second singular zig-zag line starts in the point marked 4 and intersects the dislocation of the square net in between node points. No segments of $a[100]$ dislocation are formed in such a case, but the configuration shown in Fig. 15(b) results. The zig-zag line apparently ends also at the point where the two curvilinear hexagones are visible. Looking more closely one finds that the zig-zag line in 3 is exactly the prolongation of the zig-zag in 4 and similar relations can be found between other lines of the pattern. From these we can conclude that the original configuration of lines was something like shown on Fig. 15(b). The singular lines having the same Burgers vector, annihilated mutually where intersecting, and this has given rise to the elongated curvilinear hexagones. These are in fact « holes » in the net in the sense described by FRANK [19]. The curvilinear shape of the longer side of the hexagones, results from the tendency of maintaining angles of 120° at the node points.

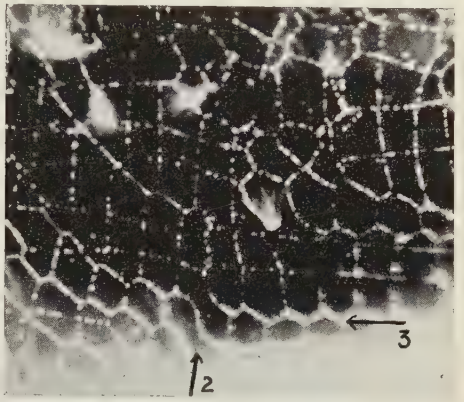


Photo 8. – Network containing zig-zag lines and « holes ». See also Fig. 15.

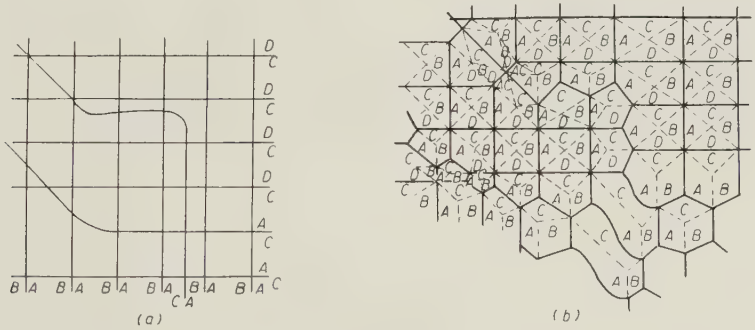


Fig. 15. – a) Configuration of lines giving rise to pattern b), which can be compared with photograph 8.

6·2. *Climb phenomena.* – Climb in ionic crystals evidently implies the emission or the absorption of numbers of vacancies or interstitials of both kinds in their stechiometric proportion. The two planes have to lengthen or to shorten simultaneously or alternatively in order to maintain charge equi-

librium. If diffusion takes place by means of pairs the first condition is automatically satisfied for the NaCl and the CsBr structure.

The decoration methods have made possible a number of interesting observations concerning climb phenomena, which may also have some bearing on the origin of dislocations in melt grown crystals. Dislocation lines having a helical shape have been observed in CaF_2 [8, 9, 32] and in NaCl [5, 33]. The axis of the helix is in all cases, within experimental error, parallel to a possible Burgers vector. Sequences of prismatic loops are sometimes associated with the helical dislocations. In CaF_2 they are in general situated on the same cylindrical surface as the helix itself (see photograph 9). In NaCl [5] they

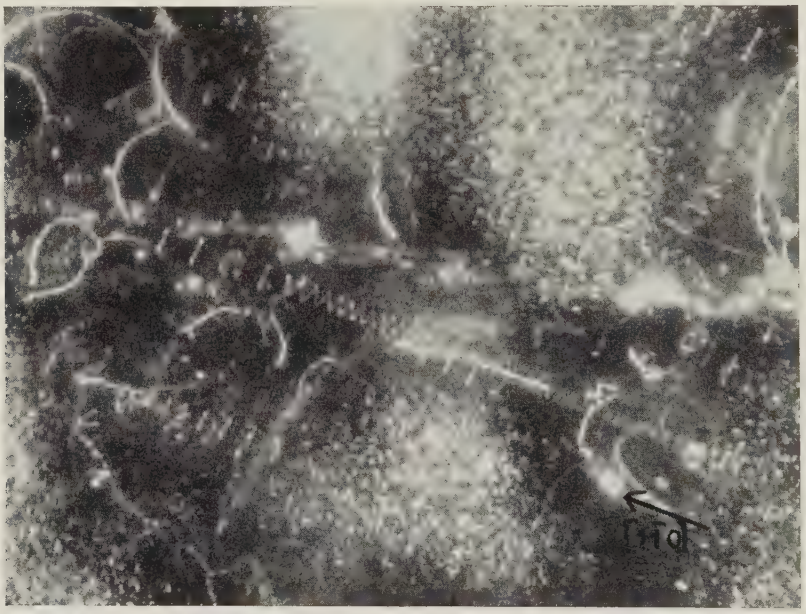


Photo 9. – Helices in CaF_2 at small magnification (after ref. [8, 9]).

occur more or less independently of the helices; they are situated on a double cone, whose top seems to be a cavity, and whose axis is again parallel to a Burgers vector. Whereas the helical dislocations predominate in CaF_2, the systems of loops seem to be more numerous than the helices in the crystals of NaCl studied by Barber *et al.*

There is agreement that these features have to be interpreted as being due to climb phenomena accompanied by prismatic glide. Two different explanations have however been proposed:

(i) The first theory [32] considers the helix as the primary phenomenon and the loops as secondary. According to this explanation the helix is ge-

nerated when a dislocation of nearly pure screw character climbs. The mechanism is related to the mechanism for the formation of spiral prismatic dislocations described by SEITZ [17]. The main difference is that the repulsion between successive turns is taken into account, so that instead of the formation of a double conical cavity, which is equivalent to a helix with a step equal to one atomic distance, a helix with a large step is developed as a consequence of prismatic glide. The prismatic loops which are sometimes observed would result from the intersection of the helix with a dislocation of the same Burgers vector.

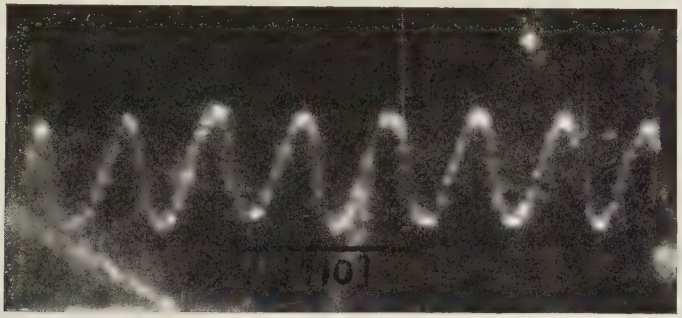


Photo 10. – Helix in CaF_2 at larger magnification (after ref. [8, 9]).

(ii) The second explanation advanced by BARBER *et al.* considers the helix as being generated from a system of loops which has been intersected by a dislocation of the same Burgers vector. The loops themselves would be generated, starting from the central cavity by the Bardeen-Herring mechanism [34].

MITCHELL (private communication) believes that in the case of the NaCl crystals, studied by his collaborators, the loop systems may have their origin in the decoration process (thermal decomposition of gold halide). In the case of CaF_2 it could be shown that helices and sequences of prismatic loops are present in crystals « as grown ». This was deduced from the presence of two kinds of etch patterns on (111) cleavage faces.

(i) Double rows of etch pits having a [110] direction, which on continued etching approach one another, coalesce and finally become flat bottomed. They are evidently due to helices or sequences of loops which have their axis parallel to the plane of observation (Photograph 11).

(ii) Regularly terraced pits of the kind reproduced in Photograph 12. It was shown that they are due to helices parallel to a [110] direction which intersects the (111) plane [35]. The step height of the terraces is related to the step of the helix.

The fact that helices and sequences of loops are present in crystals as grown, suggests strongly that their origin lies in the precipitation of vacancies, eventually combined with deformations due to thermal stresses.

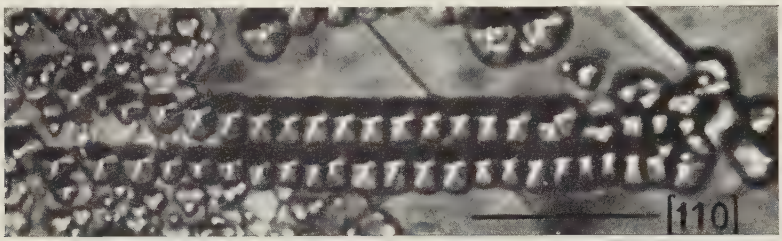


Photo 11. - Etch pattern on (111) plane of CaF_2, as grown, due to a helix having its axis parallel to (111) (after ref. [33]).

A mechanism based on the precipitation of vacancies, and which would account for the production of loop systems of the kind observed in CaF_2 and NaCl was suggested independently of the observations of BARBER *et al.* [5], and will be summarized here [36].

Suppose that during precipitation of vacancies a first dislocation ring is formed in a certain area. In a certain volume around this, the probability for the nucleation of a new ring is small, so that we can divide schematically the crystal into small regions containing each one ring at the start. The size of these regions will depend on the supersaturation at the moment of nucleation of the rings. Thermal stresses will deform some of these rings by prismatic glide, in such a way that some screw parts are formed, as e.g. shown schematically in Fig. 16(*b*). The ring will now climb further as more vacancies precipitate; the screw parts will hereby adopt a helical shape (as described previously [8, 32]). The ring will thus acquire the shape shown in Fig. 16(*c*). Further climb, combined with prismatic glide will ultimately lead to intersection of the two helices, giving

Photo 12. – Etch pattern on (111) plane of CaF_2, as grown, due to a helix having its axis inclined with respect to the observation plane (after ref. [35]).

rise to closed loops, all of the same Burgers vector, in the way shown in Fig. 16(*d*). As a consequence of repulsion between dislocations of the same Burgers vector loops will in general be « blown off » in the direction of the Burgers vector by prismatic glide and one obtains a situation as shown in Fig. 16(*d*). It is clear that this mechanism is similar to the mechanism postulated by BARDEEN and HERRING [34]. It has however the virtue that it avoids the difficulty inherent to the latter mechanism, namely to necessitate the addition or substraction of extra material between successive lattice planes. It is clear that the dislocation « sources » just described can provide in a restricted area a sufficient number of dislocations of the same kind to allow the formation of fairly regular networks. If some rings fail to develop into a « double cone source » because they remained purely prismatic, they may nevertheless be taken up into a net and become singular dislocations, or they may interact with other single loops and form a part of the three-dimensional network. It is clear that the same mechanism could originate also from a glide loop formed at a cavity; this could be the case in the crystals studied by BARBER *et al.* In the case of a cavity the source need not be double, a simple cone may result. MITCHELL (private communication) has independently thought of a similar mechanism.

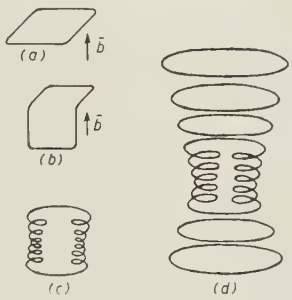


Fig. 16. – The generation of loops of prismatic dislocation as a consequence of the precipitation of vacancies in a melt grown crystal.

There is further evidence supporting this assumption.

« Double cone sources » and helices on one hand, and networks on the other hand, are usually not found in the same crystal. Crystals which contain helices and sequences of loops, as well in CaF_2 as in NaCl only rarely contain well developped networks. Vice versa the crystals of NaCl and KCl studied for networks, contain only isolated loops and rarely helices. In the case of NaCl this appears to be correlated with the way the crystals were grown. Our crystals containing networks were grown using a Czochralski method, in a small gradient, and were cooled slowly. The crystals containing the double cone sources on the other hand, were grown with a Bridgman method, in a large and steep gradient (MITCHELL, private communication). In the first case as a consequence of the relatively small supersaturation of vacancies, only few sources are nucleated per unit volume. Each of them has however a large number of vacancies at its disposal and can develop, whilst cooling, many and large loops, which on intersecting will form regular nets. The regularity of the observed nets would be difficult to explain on the basis of a random precipitation of loops.

In the second case, the large and steep gradient, will cause a high supersaturation of vacancies; a large number of sources will be nucleated per unit volume. Every one of them has however only a limited supply of vacancies

Photo 13. – Etched cube face of sodium chloride. Note that dislocations in slip traces parallel to [110] are piled up behind the intersection points with slip traces [1$\bar{1}$0].

to grow further and develop loops; they will remain in the embryonic stage; very few will be able to produce large loops. The production of cavities is presumably also a consequence of the high supersaturation of vacancies. The cavities may become unstable in some later stage; they can then transform into rings, by means of the mechanism described, by the distillation of vacancies from the cavity towards the rings. In the CaF_2 the supersaturation seems to have been somewhat smaller.

Climb phenomena have been studied also by means of etching techniques in LiF. Gilman and Johnston [6] found that on anneal glide bands breaks up and dislocations are displaced perpendicular to the glide bands. This is in fact the first stage in polygonization. The same has been observed by means of decoration methods in NaCl.

6·3. *Glide phenomena.* – We will discuss here only results obtained by methods involving direct observation.

Although the decoration methods discussed before are in principle suited to reveal dislocations immediately after deformation, and hence make deductions concerning the deformation mechanism, there is a practical limitation in that it is necessary to give a certain amount of anneal to the crystals (at the exception of the silver halides) in order to decorate the dislocations. This may disturb their configuration. To minimize the disturbance one looks for an impurity which diffuses rapidly by means of interstitial sites, already at a temperature somewhat lower than the temperature where climb begins (and preferentially lower), and which precipitates, on lowering the temperature as small but observable particles. This practically limits the choice to ions which are easily reduced and form metallic colloids. Silver in KCl [30] and gold in NaCl and KBr [5] probably satisfy this requirement.

Nevertheless striking results comparable to those obtained by DASH [37] in silicon, have not been found as yet in the alkali halides. The silver halides were studied by MITCHELL [29]: owing to the softness of the crystals and the occurrence of pencil glide, the patterns are very complicated and not easily interpreted.

On the other hand it was pointed out [22] that glide phenomena and the motion of individual dislocations could be studied by means of etch pits. Some results were obtained in the case of NaCl [22], but the method has been used with considerable success for LiF [6]. We will summarize some significant results of these experiments.

(i) Work on sodium chloride. Piling up of dislocations was found at the intersection point of glide traces (110) and $(1\bar{1}0)$ as observed in the (001) plane but not at the intersection points of (110) or $(1\bar{1}0)$ and (101). This can for instance be deduced from photograph 60 of ref. [26]; reproduced here as photograph 13. It is clearly visible that the density of etch pits diminishes suddenly after intersection points of the first kind, but not at those of the second type. We believe that the piling up is due to the formation of an a[100] dislocation at the intersection line. Although this is not a real sessile dislocation, it would in any case be much less mobile under the influence of the shear stress that made move the first two dislocations.

From the study of nets could be concluded that this kind of dislocation is stable, at least in the screw orientation. In this case it has however the edge orientation, of which the stability is doubtful according to theory [10]. The applied stress will however contribute to its formation. At the other intersections no such dislocation could form.

(ii) The work on lithium fluoride. GILMAN and JOHNSTON [6] have developed an etching and polishing technique for LiF, of which they prove that it reveals all dislocations (see characteristics in Table IV). They also worked out a method for introducing small fresh dislocation half loops

into the crystal. This is done by rolling a steel ball over the surface, and polishing away a layer of material such that only the deepest half loops remain. It is found that on subsequent stressing of the crystal, glide bands develop in the plane of the small loops, proving that some multiplication process starting from the small loop is operating. A more detailed study of the multiplication process reveals that the loop is only capable of creating the conditions which will allow multiplication, if it has been expanded to a size exceeding some critical size, which appears to be of the same order of magnitude as the average distance between dislocation in LiF. Loops which were made to collapse by reversing the stresses, after having been expanded to a smaller size, did not give rise to glide bands. It was also found that in the glide plane, in which a loop had been expanded above the critical size, loops were left behind by the moving dislocation. All these observations seem to support the idea that dislocation interactions are the sources of the defects responsible for dislocation nucleation.

The same authors also made estimates of the lattice resistance to glide of the dislocation; they find $4 \cdot 10^7$ dyn/cm^2, the stress needed to expand a half loop being $6 \cdot 10^7$ dyn/cm^2 and the stress nedeed to contract it being $2 \cdot 10^7$ dyn/cm^2. The stress needed to nucleate dislocation loops in apparently dislocation free material is found to be about 1 kg/mm^2.

REFERENCES

[1] J. N. Hedges and J. W. Mitchell: *Phil. Mag.*, **44**, 223 (1953).
[2] S. Amelinckx: *Phil. Mag.*, **1**, 269 (1956).
[3] S. Amelinckx: *Lake Placid Conference on Dislocations.*
[4] S. Amelinckx, W. Van der Vorst, R. Gevers and W. Dekeyser: *Phil. Mag.*, **46**, 450 (1955).
[5] D. J. Barber, K. H. Harvey and J. W. Mitchell: *Phil. Mag.*, **2**, 704 (1957).
[6] J. J. Gilman and W. G. Johnston: *General Electric Research Reports*, 56-RL-1560, 1620 and 57-RL-1662, also *Lake Placid Conference on Dislocations.*
[7] W. Bontinck and W. Dekeyser: *Physica*, **22**, 595 (1956).
[8] W. Bontinck and S. Amelinckx: *Phil. Mag.*, **2**, 94 (1957).
[9] W. Bontinck: *Phil. Mag.*, **2**, 561 (1957).
[10] H. B. Huntington, J. E. Dickey and R. Thomson: *Phys. Rev.*, **100**, 1117 (1955).
[11] E. Schmidt and W. Boas: *Plasticity of Crystals* (London, 1950).
[12] F. Seitz: *Rev. Mod. Phys.*, **23**, 328 (1951).
[13] A. Seeger: *Report of Bristol Conference*, The Phys. Soc. London, 1954, p. 391.

[14] H. Kanzaki: *Journ. Phys. Soc. Japan*, **11**, 120, 126 (1956).
[15] H. Kanzaki: *Journ. Phys. Soc. Japan*, **10**, 663 (1955).
[16] Z. Guylai and D. Hartly: *Zeits. f. Phys.*, **51**, 378 (1928); **78**, 630 (1932).
[17] F. Seitz: *Advances in Physics*, **1**, 43 (1952).
[18] W. G. Johnston: *Phys. Rev.*, **98**, 1877 (1955).
[19] F. C. Frank: *Report of Bristol Conference*, The Phys. Soc. London, p. 159 (1955).
[20] S. Amelinckx: *Physica* (in the press).
[21] F. C. Frank: *Carnegie Institute of Technology, Symposium on the Plastic Deformation of Cristalline Solids* (*Pittsburgh Report*), p. 150.
[22] S. Amelinckx: *Acta Metall.*, **2**, 848 (1954).
[23] F. Seitz: *Rev. Mod. Phys.*, **26**, 7 (1953).
[24] J. Okada: *Journ. Phys. Soc. Japan*, **10**, 1018 (1955).
[25] S. Amelinckx and E. Votava: *Naturwiss.*, **41**, 422 (1954).
[26] S. Amelinckx: *Natuurwetensch. Tijdschr.*, **36**, 4 (1954).
[27] J. J. Gilman: *Transactions AIME*, 1252 (1955).
[28] J. W. Mitchell: private communication.
[29] J. W. Mitchell: *Lake Placid Conference on Dislocations.*
[30] S. Amelinckx: *Acta Metall.*, in the press (1957).
[31] W. Maenhout-Van der Vorst and W. Dekeyser: *Phil. Mag.*, **1**, 882 (1956).
[32] S. Amelinckx, W. Bontinck, W. Dekeyser and F. Seitz: *Phil. Mag.*, **2**, 355 (1957).
[33] S. Amelinckx, W. Bontinck and W. Maenhout-Van der Vorst: *Physica*, **23**, 270 (1957).
[34] J. Bardeen and C. Herring: *Imperfections in Nearly Perfect Crystals* (New York, 1952), p. 261.
[35] S. Amelinckx, W. Bontinck and W. Dekeyser: *Phil. Mag.* (in the press).
[36] S. Amelinckx and W. Bontinck: *Acta Metall.*, **5**, 345 (1957).
[37] W. Dash: *Journ. Appl. Phys.*, **27**, 1193 (1956).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1957
1° Trimestre

Mobility of Electrons in the Silver Halides.

F. C. Brown

University of Illinois - Urbana, Ill.

1. – Introduction.

The silver halides are convenient materials in which to study the interaction between conduction electrons and the ionic lattice. In a polar crystal such as AgCl it is believed that electrons are scattered mainly by the longitudinal optical mode of vibration of the lattice rather than by the acoustical modes. The strength of the electron-lattice interaction is fairly high, high compared to a material such as PbS for example, where the interaction is weakened by electronic polarization [2]. Theoretical calculations of the temperature dependence of mobility in polar crystals have been given by Low and Pines [3] and they emphasize that their results are applicable at temperatures below the Debye temperature corresponding to the longitudinal optical mode, $\Theta = 280$ °K for AgCl.

An unusual property of AgCl and AgBr allows the determination of mobility by the measurement of the time of transit of electrons across the thickness of a crystal. It has been found that in properly prepared single crystals the mean free time before trapping, T, can be of the order of microseconds giving rise to values of schubweg per unit field $w/E = 10^{-4}$ to 10^{-3} cm²/V. This property arises because of a very low density of *deep* trapping levels ($\sim 10^{13}$ cm^{-3}) or because of small trapping cross-sections ($\sim 10^{-17}$ cm²) or both.

Recently a study has been made of the factors which influence this lifetime in AgCl by means of pulsed fields, pulsed light measurements at room temperature [4]. The long lifetimes are shown to be characteristic of crystals grown in air from a moist starting charge. On the other hand, crystals grown in vacuum, halogen or an inert atmosphere show lifetimes before trapping of 10^{-8} s or less, more like the case of the alkali-halides. Crystals of both types

have been prepared with a very low concentration of metal impurities of the order of 0.1 parts per million, as determined by spectrochemical analysis. Table I shows values of lifetimes for comparison.

TABLE I.

Substance	Lifetime (s)
AgCl (air)	10^{-5}
AgBr (air)	10^{-6}
AgCl (vacuum)	$5 \cdot 10^{-9}$
KCl	$5 \cdot 10^{-10}$

2. – Primary currents. Transit time method.

Imagine a sample of the silver halides cooled to low temperature to minimize ionic conductivity and placed between plane parallel electrodes separated by a distance l. The time constant of the crystal capacitance and preamplifier input resistance is large so that integrated charge is measured. The amplifying system is wide band so as to faithfully reproduce the time dependence of the collected charge. If n_0 electrons are released at time $t = 0$ at the cathode $X = 0$, they are swept into the volume by the electric field, $E = V/l$, until they are either trapped in the volume or collected at the anode. Recombination is usually unimportant since the holes remain on the surface adjacent to the cathode. Assuming an uniform distribution of deep traps the number of electrons free at any time t is $n = n_0 \exp[-t/T]$, where T is the mean trapping time. During a time $\mathrm{d}t$ the charge induced on the electrodes is

$$\mathrm{d}q = ne \frac{\mathrm{d}x}{l},$$

where

$$\mathrm{d}x = \mu E \, \mathrm{d}t .$$

The total charge $q(t)$ is

$$q(t) = \frac{n_0 e \mu E}{l} \int_0^t \exp\left[-\frac{t}{T}\right] \mathrm{d}t .$$

Using $w = \mu E T$ and integrating we get

$$q(t) = \frac{n_0 e w}{l} \left(1 - \exp\left[-\frac{\mu E t}{w}\right]\right), \tag{1}$$

which is applicable for $t \leqslant l/\mu E$.

For times greater than $l/\mu E$ the total charge is

$$Q(\infty) = \frac{n_0 e w}{l} \left(1 - \exp\left[-l/w\right]\right),$$

which is an expression for the saturation curves characteristic of primary photoresponse. In the case that trapping is less important so that $\omega \gg \mu E t$ equation for $g(t)$ reduces to

$$g(t) \cong \frac{n_0 e \mu E}{l} t .$$

Here the charge pulse has a linear rise until a time $\tau = l/\mu E$ corresponding to collection of the n_0 electrons. In practice this condition can be achieved over a wide voltage range in selected crystals so that one can determine mobility from the risetime of the pulse.

Recent measurements by Prof. F. DART and the author [5] have been made by means of a pulsed photon beam of less than $2.5 \cdot 10^{-8}$ s duration. The crystal 1.5 cm in diameter and 0.23 cm thick is arranged between electrodes one of which is transparent to pulses of strongly absorbed ultra violet light provided by a high speed rotating mirror (2 000 rev. per s) and a water cooled mercury arc. A helium cryostat capable of holding the sample at any desired temperature to 4 °K was designed. Since the risetime may be as short as 10^{-7} s the amplifying equipment has a response time of $1.8 \cdot 10^{-8}$ s and will allow the observation of charges as small as 10^{-14} coulombs induced across the crystal capacitance.

The transit time is usually long compared to the light pulse duration and the amplifier response time. However, the latter two quantities are constant with respect to the applied voltage whereas the transit time varies as $1/V$. In fact the observed risetime is given by

$$\tau_0 = \frac{l^2}{\mu V} + \tau_{\text{inst.}} .$$

In practice the crystal is held at constant temperature and the average τ_0 recorded for 10 to 25 pulses at each of several voltages. Mobility μ is then determined from the slope of the τ_0 versus $1/V$ plot whose slope should be constant at high voltages and which has a y intercept equal to the instrument time, $\tau_{\text{inst.}}$. This method of analysis also has the effect of separating out the time spent in shallow trapping states if these are not present in too high a concentration.

3. – Results and discussion.

Earlier experiments on mobility in AgCl were carried out down to liquid nitrogen temperature using β-rays to release charge carriers. From these mobility μ was shown to have an approximate $T^{-\frac{3}{2}}$ dependence down to 86 °K. Such a dependence would be in agreement with scattering by the acoustic modes rather than optical modes. It is also interesting to note that Hall mobility determined by the Redfield technique lies just a little above these drift mobility values at 86 °K.

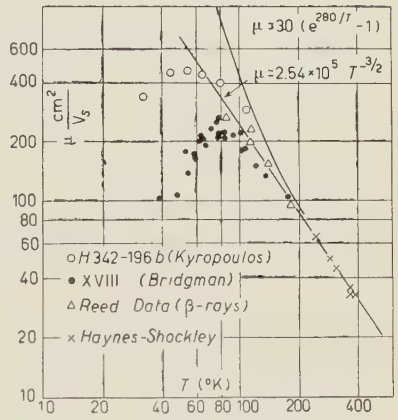


Fig. 1.

Results using pulses of light extend to very much lower temperatures and are shown in Fig. 1. Notice that a maximum in mobility occurs near 250 cm²/V s at 80 °K for one crystal grown by the Bridgman technique, whereas a maximum of about 480 cm²/V s was found at lower temperatures in the case of a crystal grown by the Kyropoulos technique. The data for this latter crystal from 65 to 110 °K lie well above the $\mu = 2.54 \cdot 10^5\, T^{-\frac{3}{2}}$ dependence shown to fit the older points (Reed data) down to 90 °K. It is not possible to obtain a good fit to the data for either crystal in the intermediate temperature range by combining any reasonable low temperature scattering law (T, $T^{\frac{3}{2}}$ or T^2) with a $T^{-\frac{3}{2}}$ dependence at higher temperatures.

Apparently a strong imperfection scattering mechanism dominates at low temperatures. If this scattering were due to ionized impurities, it would appear to require in the vicinity of 10^{17} to 10^{18} per cm³, at least in the case of the Bridgman crystal [7]. This is more than the total metal impurity content which is thought to be more nearly 10^{15} to 10^{16} per cm³. Dislocations may play a role in this scattering.

If we combine reciprocal mobilities according to

$$\frac{1}{\mu} = \frac{1}{\mu_0} + \frac{1}{\mu_i},$$

a good fit to the data for crystal H 342 can be obtained for $\mu_i = 10.5\,T$ and $\mu_0 = 30\,(\exp[280/T] - 1)$. Admittedly this method of combining inverse mobilities leads to error in the intermediate temperature range, but it is the best we can do until we know more about the energy dependence of the low tem-

perature scattering mechanism. The data strongly support a mobility theory in which the electron is scattered by optical modes of vibration. The theory of Low and Pines gives the following expression in c.g.s. units for average mobility as a function of temperature in the range $T < \Theta$:

$$\mu = \frac{1}{2\alpha\omega} \frac{e}{m} \left(\frac{m}{m^*}\right)^3 f(\alpha) \exp\left[(\Theta/T) - 1\right].$$

In this expression α is a constant which can be expressed in terms of the long wave longitudinal mode frequency $\omega = (\varepsilon/\varepsilon_0)^{\frac{1}{2}} 2\pi c/\lambda_t$ and the ratio of electron effective mass to free electron mass m/m_0. This latter quantity m/m_0 may in fact be considered an adjustable parameter of the theory. The polaron effective mass has been shown to be $m^* = m(1 + \alpha/6)$. Comparing experiment, $\mu_0 = 30\,(\exp[280/T] - 1)$, with theory we arrive at the following value of effective mass for AgCl, $m/m_0 = 0.215$, from which the polaron effective mass and interaction constant are calculated:

$$m^*/m_0 = 0.28\,,$$

$$\alpha = 1.82\,.$$

We have used the residual wave data to compute ω and from this $\Theta = \hbar\omega/h$. Uncertainties in this value of Θ probably introduce an error in the above figures somewhat greater than 10%. The effective mass we deal with here is of course averaged over direction and applies for slow electrons near the bottom of the conduction band.

REFERENCES

[1] H. Fröhlich and N. F. Mott: *Proc. Roy. Soc.* (*London*), A **171**, 496 (1939); A **62**, 446 (1949).
[2] R. L. Petritz and W. W. Scanlon: *Phys. Rev.*, **97**, 1620 (1955).
[3] F. Low and D. Pines: *Phys. Rev.*, **91**, 193 (1953); **98**, 414 (1955).
[4] F. C. Brown: to be published.
[5] F. C. Brown and F. E. Dart: *Phys. Rev.*, **108**, 281 (1957).
[6] F. C. Brown: *Phys. Rev.*, **97**, 355 (1955); **92**, 502 (1953).
[7] F. Blatt: *Journ. Phys. Chem. Solids*, **1**, 262 (1957).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Dielectric Relaxation of Lattice Defects in Crystals.

Y. Haven and J. H. van Santen

Philips Research Laboratories - Eindhoven

1. – Introduction.

The dynamical dielectric constant $\varepsilon(\omega) = \varepsilon'(\omega) - i\varepsilon''(\omega)$ of polar materials characterized by a single relaxation time τ is given by the well known Debye formulae [1]

$$\varepsilon' = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + \omega^2\tau^2} \,, \tag{1a}$$

$$\varepsilon'' = \frac{4\pi\sigma_s}{\omega} + (\varepsilon_s - \varepsilon_\infty)\frac{\omega\tau}{1 + \omega^2\tau^2} \,. \tag{1b}$$

Here ε_s and ε_∞ denote the static and high-frequency dielectric constants, σ_s is the direct current conductivity and ω stands for the angular frequency.

On the other hand, the relaxational behaviour of these materials is described by an exponential decay function $\exp[-t/\tau]$, which describes the decay of the dielectric displacement when an electric field is suddenly applied or suddenly released.

For the present purpose we will not distinguish between the (macroscopic) relaxation times τ introduced above and the microscopic relaxation times characterizing the molecular processes involved. This will be a fair approximation in systems where the dipolar contribution to the dielectric constant is only small.

Equations (1) are easily generalized for systems characterized by several relaxation times, the decay of the dielectric displacement then being given by a sum of exponentials. Each of the terms $\omega\tau/(1 + \omega^2\tau^2)$ is represented by

a curve of type A in Fig. 1 in a plot of ε'' or $\log \varepsilon''$ against $\log \omega$ (Debye curve). In such a plot, thus, one obtains a kind of absorption spectrum.

Originally equations (1) were used for discussing the dielectric properties of systems containing dipolar molecules but there is theoretical and experimental evidence that Debye's equations can also be used for describing the dielectric properties of non-ideal solids. For solids the two-potential-well model where a charged ion can move between two potential wells separated by a barrier has been discussed by FRÖHLICH [2]. BRECKENRIDGE [3] used Debye's equations for explaining the dielectric properties of lattice defects in crystals. A sodium vacancy in NaCl associated with a chlorine vacancy is a complex with dipolar properties which can orient itself in an electric field. The dielectric losses due to this type of lattice defects could not be identified with certainty but for alkali halides with divalent cation impurities losses of the Debye type were found, which were ascribed to associations of positive-ion vacancies with positive divalent ions.

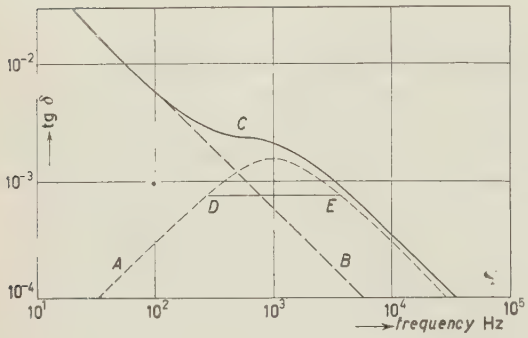


Fig. 1. – A = dielectric losses resulting from dipoles, B = conductivity losses, C = total losses, DE = half-value width of the Debye peak.

The latter type of dielectric losses has been found for NaCl containing small mounts of $CaCl_2$ or $MnCl_2$ (HAVEN [4], JACOBS [5]), for KCl containing $SrCl_2$ or $PbCl_2$ (BURSTEIN, DAVISSON and SCLAR [6]), for AgBr containing $CdBr_2$ (TELTOW and WILKE [7]) and for several other alkali halides containing divalent impurities (DRYDEN and MEAKINS [8]). Moreover, in several other substances dielectric losses were found which possibly may be ascribed to a similar type of associations between lattice defects (cf. FREYMANN and FREYMANN [9]). In these examples the experimental loss curves are all of type C in Fig. 1.

From these examples it appears that the study of the dielectric losses may be a method for studying lattice defects in crystals. In this paper we shall discuss some relations between the relaxation times and the molecular parameters characterizing the lattice defects (jump probability, spatial configuration of the defects, etc.). We shall discuss the dielectric properties of NaCl containing some $CaCl_2$ in terms of « relaxational modes », as was already indicated in a previous letter [10].

2. – Relaxational modes.

Consider a NaCl crystal with a small number of independent systems consisting of a fixed Ca^{2+} ion associated with a Na^{+}-vacancy. The vacancy is supposed to move on twelve equivalent sites round the Ca^{2+} ion. In equilibrium each of the twelve sites neighbouring the Ca^{2+} ion is occupied with equal probability by the vacancy and then the probability distribution of this system has the cubic symmetry O_h. A probability distribution deviating from the equilibrium one will decay until equilibrium has been reached. This decay can be described with the help of exponential functions of the form $\exp[-t/\tau]$, where τ is a decay or relaxation time, and a convenient description can be obtained by introducing relaxational modes [10]. By a relaxational mode we understand a probability distribution (for the vacancy in the present case), which on all sites decays with one and the same relaxation time. It can be shown, that, for a given finite system, a finite number of relaxation modes suffices to describe any decay.

There is a close correspondence between relaxational modes and the vibrational modes as introduced for the description of the vibrations of atoms in molecules [11]. The two types of modes are compared in Table I.

TABLE I. – *Comparison of relaxation and vibrational modes.*

	Relaxation	Vibration
Behaviour of free system as function of time	Decay function $\exp[-t/\tau]$	Periodic function $\exp[-i\omega_0 t]$
Describing deviation from	Equilibrium distribution (probability)	Equilibrium positions (spatial)
Number of sites or atoms	N	N
Number of modes « degree of freedom »	N	$3N$

For the more simple systems the relaxational modes are easily derived and it is not even necessary to introduce the concept of mode. For the more complicated systems, however, especially of high symmetry, it is useful to discuss the dielectric properties in terms of relaxational modes.

In the case of high symmetry the results of group theory may be used. The relaxational modes correspond with the irreducible representations of the

corresponding symmetry group. For instance, the changes in the probability distributions brought about by electric fields are described by modes that belong to the same irreducible representations as do polar vectors (cf. also FALK and MEIXNER [12]). These modes are called electrically active modes.

3. – Relaxational modes of the (Ca^{2+}-Na vacancy) system in NaCl.

The relaxational modes of the system of a fixed Ca^{2+} ion in NaCl with a Na-vacancy in one of the twelve neighbouring Na-sites (symmetry group O_h) are shown in Fig. 2. The Na-vacancy is assumed to jump from a site to a

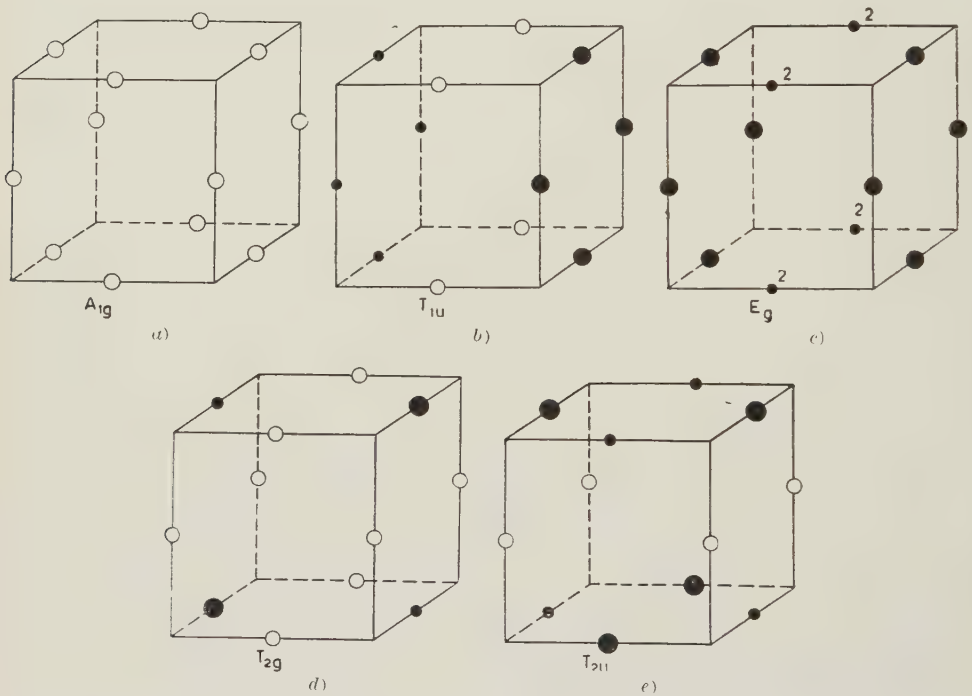


Fig. 2. – The five relaxation modes of the Ca^{2+}+vacancy system in NaCl. The Ca^{2+} is in the center of the cube; the Na-vacancy is distributed over the twelve midpoints of the edges. ○, ● and ● denote distributions that are equal to, larger or smaller than the equilibrium distributions, respectively.

given neighbouring site with a jump probability k (thus making $4\,k$ jumps per second).

The mode A_{1g} (Fig. 2*a*) is the equilibrium distribution, where the vacancy

is with equal probability on any of the twelve sites. The relaxation time is given by $1/\tau = 0$.

The three-fold degenerate mode T_{1u} (Fig. 2*b*) is an electrically active relaxational mode, which will be excited by application of electric fields. Its relaxation time is $\tau = 1/2k$, so that its maximum energy absorption lies at a frequency $f = 2k/2\pi$ of the applied electric field.

The two-fold degenerate mode E_g (Fig. 2*c*) as well as the three-fold degenerate mode T_{2g} (Fig. 2*d*) are the so-called mechanical relaxational modes. The mode E_g will be excited by the application of a unidirectional stress along the *c*-axis (or *a*- or *b*-axes) and will have its maximum energy absorption at a frequency $f = 6k/2\pi$ of the applied stress. The mode T_{2g} will be excited by application of shear stresses and has its maximum energy absorption at a frequency $f = 4k/2\pi$ of the applied stress.

The mode T_{2u} (Fig. 2*c*) is of a more complicated nature, with a relaxation time $\tau = 1/6k$. There is no method for exciting this relaxational mode with the aid of homogeneous fields or stresses.

The results of Fig. 2 show that there is in this system only one electrically active, but two mechanically active relaxational modes. In the present example the relaxation times are in the proportion $\tau(T_{1u}):\tau(E_g):\tau(T_{2g}) = 1:\frac{1}{3}:\frac{1}{2}$. We are not aware of examples, where the dielectric and mechanical relaxation have been measured on the same sample (cf., however, BRECKENRIDGE [3*b*]). To a certain extent some published measurements on quartz crystals can be used for such a comparison. VOLGER, STEVELS and VAN AMERONGEN [13] found in quartz (plate parallel to the major rhombohedral face; normal at 52° to the *c*-axis) a dielectric relaxation peak at 32 kHz and 38.5 °K, whereas an extrapolation of mechanical loss measurements on quartz (AT cut normal at 55° to the *c*-axis) by BÖMMEL, MASON and WARNER [14] gives a relaxation peak at 93 kHz and 38.5 °K. Thus the two relaxation times are in the proportion $\tau_{\text{electr}}:\tau_{\text{mech}} = 3:1$ (subject to a rather large uncertainty).

So far we had assumed the Ca^{2+} ion in the NaCl-$CaCl_2$ system to be in a fixed position. Removing this restriction, one can notice, firstly, that neither the number of sites round Ca^{2+} available for the vacancies is altered nor their symmetry properties. Therefore, the relaxational modes are the same whether the Ca^{2+} is assumed to be mobile or not. If the Ca^{2+} ion jumps with a probability k_{Ca} from one site to a given other, one derives from Fig. 2*a* the relaxation frequency $1/\tau$ to be $1/\tau = (2k + 2k_{Ca})$ for the electrically active mode T_{1u} (Fig. 2*b*). The jump frequency of the Ca^{2+} ion, however, does not enter into the expressions for the relaxation times of the two mechanically active modes E_g and T_{2g} (Fig. 2*b* and 2*d*). Therefore, an accurate measurement of the electrical and mechanical relaxation times may make possible a separation of the jump frequencies of the Ca^{2+} and Na^{+} ions in an associated complex. The relaxation time for the T_{2u} mode (Fig. 2*c*) will be $1/\tau = (6k + 2k_{Ca})$.

4. – More refined dipole-centre models.

If the Ca^{2+} ion is taken at site (000) and the twelve neighbouring Na-sites are (110), etc., then the next nearest sites are the six sites (200) etc. Let the six sites (200) be accessible to the Na vacancy with equal probability as the twelve (110) sites. Then, the number of degrees of freedom is $12+6=18$, so that one has to look for six additional relaxational modes.

One finds the equilibrium mode A_{1g}, as before, and a second relaxational mode A_{1g} which represents a distribution (p_0+p_1) on all sites (110) and (p_0-2p_1) on all sites (200), where p_0 denotes the equilibrium distribution. The latter mode has a relaxation time $\tau=1/2k$ and it can be excited by hydrostatic pressure.

There are now two three-fold degenerate electrically active modes T_{1u} with relaxation times $\tau=1/2k$ and $\tau=1/6k$. In the first mode the probability distribution is (p_0+p_1) on (110) and (p_0+2p_1) on (200), with (p_0-p_1) on $(\bar{1}10)$ and (p_0-2p_1) on $(\bar{2}00)$, etc. In the second mode the distribution on the four sites mentioned above is (p_0+p_1), (p_0-2p_1), (p_0-p_1), and (p_0+2p_1), respectively. The electric moment connected with the latter mode, however, happens to be zero, so that also in this case only one electrical relaxational mode plays a part.

Furthermore, there are two two-fold degenerate modes E_g, which may be excited by unidirectional stresses (along a-, b-, or c-axes). The relaxation times are $\tau=1/(6+\sqrt{6})k$ and $\tau=1/(6-\sqrt{6})k$.

The two modes T_{2g} and T_{2u} found in the previous section will not be changed and not be doubled by the addition of the (200)-sites. The relaxation times of the two modes are $\tau=1/6k$ and $\tau=1/8k$, respectively. In these modes the sites (200) keep their equilibrium probability distributions.

The case where the sites (200) and (110) have different equilibrium occupation probabilities will not be considered in detail and we give only some results pertaining to the two electrically active relaxation modes T_{1u}.

Let the probability for a jump between neighbouring (110)-sites be k in either direction, and let for a jump from an (110)-site to a neighbouring (200)-site the probability be kp, and k/p in the opposite direction. The equilibrium distribution on a (200)-site will then be p^2 times that on an (110)-site.

The relaxation times, obtained by solving the secular equation, are given by

$$1/\tau_{\pm}=k[2+p+p^2\pm\{(2+p+p^2)^2-8p-4p^2\}^{\frac{1}{2}}]/p$$

or if $p\ll 1$

$$1/\tau_{+}=(4/p+p)k\,,$$

$$1/\tau_{-}=(2+p)k\,.$$

If the ratio of the excess of the probability distribution on a (200)-site and the excess on an (110)-site is denoted by R, we find for the two roots

$$R_{\pm} = -2 + p + p^2 \mp \{(2 + p + p^2)^2 - 8p - 4p^2\}^{\frac{1}{2}},$$

or if p is small

$$R_{+} = -4 + 2p + p^2, \qquad R_{-} = p^2.$$

If an electric field is applied the contributions to the polarization of the (+) and (—) mode are in a proportion of $\frac{1}{4}p^2$:1, respectively (for small values of p).

These results show that by application of an electric field mainly one of the two electrically active relaxational modes will be excited, the frequency having been shifted slightly from the relaxation frequency obtained in the preceding section.

For a value of $p = 0.5$ the contribution to the polarization of this second mode is about 7 % of the excitation probability of the first one, the equilibrium distribution on (200) being about 0.25 of that on (110).

From this result one may conclude that it will be difficult to draw conclusions from dielectric loss measurements, as to the detailed structure of the dipole centres.

REFERENCES

[1] P. Debye: *Polare Molekeln* (Leipzig, 1929).
[2] H. Fröhlich: *Theory of Dielectrics* (Oxford, 1949).
[3] R. G. Breckenridge: *a) Journ. Chem. Phys.*, **16**, 959 (1948); **18**, 913 (1950); *b) Imperfections in Nearly Perfect Crystals* (New York, 1952), p. 219; T. B. Grimley: *Journ. Chem. Phys.*, **16**, 620 (1949).
[4] Y. Haven: *Journ. Chem. Phys.*, **21**, 171 (1953); *Rep. Confer. on Defects in Crystalline Solids (Bristol, 1954)*, p. 261 (1955).
[5] G. Jacobs: *Naturw.*, **42**, 575 (1955).
[6] E. Burstein, J. W. Davisson and N. Sclar: *Phys. Rev.*, **96**, 819 (1954).
[7] J. Teltow and G. Wilke: *Naturw.*, **41**, 423 (1954).
[8] J. S. Dryden and R. J. Meakins: *Discussions Faraday Society* **23**, 39 (1957).
[9] M. Freymann and R.Freymann: *Journ. Chem. Phys.*, **20**, 1970 (1952).
[10] Y. Haven and J. H. van Santen: *Journ. Chem. Phys.*, **22**, 1146 (1954).
[11] G. Herzberg: *Infrared and Raman spectra of Polyatomic Molecules* (New York, 1945).
[12] G. Falk and J. Meixner: *Zeits. f. Naturf.*, **11***A*, 782 (1956).
[13] J. Volger, J. M. Stevels and C. van Amerongen: *Philips Res. Rep.*, **10**, 260 (1955).
[14] H. E. Bömmel, W. P. Mason and A. W. Warner: *Phys. Rev.*, **102**, 64 (1956).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

The Electronic Structure of a V-Center.

T. G. CASTNER

University of Illinois - Urbana, Ill.

W. KÄNZIG and T. O. WOODRUFF

General Electric Research Laboratory - Schenectady. N.Y.

1. – Introduction.

If alkali halide crystals are irradiated with X-rays, many types of light absorbing centers, so called color centers, are produced. The relative abundance and the production rates of the various types of centers depend upon the conditions of irradiation (e.g. dose, wavelength, crystal temperature) as well as upon the imperfections of the crystal (impurities, vacancies and interstitials, dislocations). The detailed structure of most of these color centers is not known, because a powerful experimental tool was lacking for a long time. However, plausible models were deduced from the properties of the crystal containing color centers. For example, a model for the F-center was proposed by DEBOER as early as 1937 [1]. It consists of a halide vacancy at which an electron is trapped. It was not until 1953 that this model was confirmed through the electron spin resonance experiments by KIP, KITTEL, LEVY and PORTIS [2]. It was shown that the trapped electron is shared, essentially as a valence electron, by the six alkali ions surrounding the halide vacancy.

The F'-center is another color center the nature of which was recognized early. PICK demonstrated experimentally that it consists of an F-center that has trapped an electron [3].

Models for other electron excess centers (F-type centers) were proposed by SEITZ [4]. They all consist of simple vacancy aggregates that have trapped one or two electrons. Direct experimental confirmation of these models is still lacking.

The presence of electron excess centers in X-rayed crystals implies the existence of electron deficient centers. These are called V-centers because their optical absorption was found on the violet side of the absorption bands of the electron excess centers. SEITZ [4] proposed that the V-centers are the « antimorphs » of the F-type centers. Their models are obtained from the models of the F-type centers by replacing the trapped electrons by holes and the halide vacancies by alkali vacancies and vice versa. These models are rather speculative because not enough experimental work has been done. Also, it is difficult to see why a hole, an electron deficiency in the otherwise filled shell of a halide ion, should behave like an excess electron, attached as a valence electron to the closed shell of an alkali ion.

Conventional paramagnetic resonance techniques can be applied very successfully to the study of V-centers because the hyperfine interaction of the incomplete electron shell with the halogen nucleus is rather large. Furthermore, it turns out that trapped holes can be much more localized than trapped electrons so that the hyperfine splitting is readily resolved. In this paper we summarize the main results of the structure determination of a V-center by means of paramagnetic resonance. A detailed description and analysis of the spectra has been published elsewhere [5, 6].

The V-center to be discussed here was present in pure synthetic alkali halide crystals (*) KCl, NaCl, LiF and KBr upon X-irradiation at 80 °K. In KCl and KBr it anneals out roughly in the same temperature region where the optical V_1-band bleaches [7]. This might suggest a correlation between the two processes (+). However, this point requires further experimental investigation since we did not measure simultaneously the optical absorption and the paramagnetic resonance absorption (×). The concentration of these V-centers is roughly of the same order of magnitude as the concentration of the F-centers created simultaneously. In the cases of LiF and KCl we convinced ourselves that the V-center is also present upon X-irradiation at 20 °K.

The X-rayed crystals contain several different types of electron excess and electron deficient centers that are paramagnetic. However, in most cases the conditions of the experiment can be chosen so that the observation of the paramagnetic resonance of a particular type is at least strongly favored. For example, the paramagnetic relaxation times of the various types differ substantially. Therefore, it is possible to choose a temperature and a radio-

(*) Obtained from the Harshaw Chemical Company, Cleveland, Ohio.

(+) In the early stages of this work we believed to observe the paramagnetic resonance of the V_1-center [8].

(×) *Note added in proof.* Recently DELBECQ and YUSTER showed that this V-center is not identical with the V_1-center. It is a *new* center, which gives rise to an optical absorption band that is close to the V_1-band (private communication).

frequency power level such that « undesirable » resonances do not show up, either because of complete saturation or excessive broadening. The resonance absorption of the V-centers under consideration is readily observed at 80 °K with a microwave magnetic field H_1 not exceeding 1 gauss.

2. – The results of the analysis of the paramagnetic resonance spectra.

2·1. *The model of the V-center.* – The study of the hyperfine splitting clearly showed that the paramagnetic resonance was due to an electron deficiency (hole) equally shared by two equivalent halide ions. This imperfection is best described as a halogen$_2^-$ molecule-ion. To a good approximation it is axially symmetric. The axis (i.e. the line joining the two nuclei) is *exactly* a [110] axis of the crystal. The six [110] axes of the crystal are equally populated with molecule-ions which may be regarded as frozen in at 80 °K. We expect, however, that jumps from one halide pair to another may occur if the molecule-ion is optically excited. In the more detailed paper [5] evidence has been presented that the molecule-ion probably is not associated with vacancies or impurities but possibly might be regarded as a self-trapped hole. Fig. 1 schematically depicts the position of the molecule-ion in the crystal lattice. It seems likely that the hole never migrated through the lattice, but stayed trapped near or at the site where it was created.

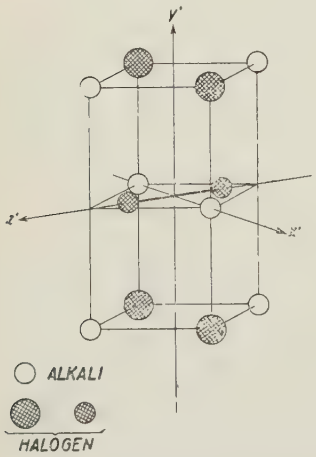


Fig. 1. – Orientation and principal axes of the halogen$_2^-$ molecule-ions in the alkali halide crystals. The molecular bond is symbolized by a heavy line.

2·2. *The spin Hamiltonian.* – The hyperfine splitting in the paramagnetic resonance spectra depends upon the angle θ between the axis z' of the molecule-ion and the magnetic field; it decreases with increasing θ. Since axial symmetry is a good approximation the following spin Hamiltonian gives an adequate description of the spectra

$$\frac{\mathcal{H}}{g_0\beta_0} = \frac{g}{g_0} H_0 S_z + a\boldsymbol{I}\cdot\boldsymbol{S} + bI_{z'}S_{z'} ,$$

where $\beta_0 =$ Bohr magneton,

$g_0 =$ spectroscopic splitting factor of free electron,

g = spectroscopic splitting factor of molecule-ion for orientation visualized (very close to g_0),

z = axis of H_0,

z' = molecular axis,

$\boldsymbol{I} = \boldsymbol{I}_1 + \boldsymbol{I}_2$, with $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ the nuclear spin operators of the two nuclei in the molecule-ion.

The transitions resulting from this spin Hamiltonian were calculated and the constants a and b chosen to fit the experimental spectra. Their values are listed in Table I. For angles θ below 75° second-order perturbation theory is sufficiently accurate, whereas for $\theta > 75°$ a more rigorous treatment is necessary [6]. The sign of the constant a could not be determined by these calculations; other considerations led us to favor the positive sign.

TABLE I. – *Constants of the spin Hamiltonian.*

(The sign of a is undetermined in the analysis of the spin Hamiltonian. Other considerations, however, seem to favor a positive value [5]).

Crystal	Isotopes	$(a+b)$ gauss	$\lvert a \rvert$ gauss
LiF	$^{19}F - ^{19}F$	887	59
KCl	$^{35}Cl - ^{35}Cl$	101	9
NaCl	$^{35}Cl - ^{35}Cl$	98	9
KBr	$^{81}Br - ^{81}Br$	455	80

3. – The wave function of the unpaired electron.

In the LCAO approximation the orbital ground state of the unpaired electron of the halogen$_2^-$ molecule-ion is a linear combination of two atomic orbitals φ_1 and φ_2 each centered on one of the two nuclei of the molecule-ion. The resulting orbital may be considered as an antibonding orbital since it weakens the bond of the halogen$_2$ molecule. The atomic orbitals φ_1 and φ_2 are primarily of p-character because the molecule-ion may be thought of as created by removing a p-electron from the noble gas configuration of the halide ions ($2p$, $3p$ and $4p$ for the fluorides, chlorides and bromides respectively). The degeneracy of these p-functions is lifted through the formation of molecular orbitals. The ground state corresponds to a $p_{z'}$ function (z' designates the molecular axis). This means that the orbital moment is *quenched*. The anisotropy of the hyperfine splitting arises from the classical interaction between the electron magnetic moment and the nuclear magnetic moments.

A considerable admixture of s-type function has to be expected because the two halide ions involved in the molecular bond polarize each other electrically. We assumed that the s-type function admixed to the np-function is a $(n+1)s$-function and calculated the fractions α^2 and β^2 of p and s-character from the observed anisotropy of the hyperfine interaction (from the constants a and b of the spin Hamiltonian). The results are listed in Table II.

TABLE II. – *Fraction of p and s-character of the atomic orbitals φ, as calculated from the values of the constants « a » and « b » of the spin Hamiltonian. The result is given for both signs of the constant « a ». (The positive sign is more likely to be the correct choice).*

Crystal	Sign of of a	Fraction of p-character	Fraction of s-character
LiF	+	0.49	0.51
	−	0.59	0.41
KCl	+	0.62	0.38
	−	0.74	0.26
NaCl	+	0.62	0.38
	−	0.75	0.25
KBr	+	0.57	0.43
	−	0.72	0.28

4. – The excited states of the halogen$_2^-$ molecule-ion.

The observed spectroscopic splitting factor deviates slightly from the free electron value. Table III gives the measured g-tensor referred to the axes $x'y'z'$ of the molecule-ion (see Fig. 1). The component $g_{z'}$, is very close to the free electron value $g_0 = 2.002\,29$. The components $g_{x'}$ and $g_{y'}$ show a substantial *positive* deviation from g_0, indicating that the resonance arises from an electron *deficiency* in the noble gas configuration of the halide ions.

In terms of perturbation theory the g-shift is a second-order effect and hence involves the admixture of orbitally excited states. In order to find these we first build up a « pseudomolecule » consisting of two halide ions. There are twelve np electrons to be accommodated. The following levels have to be expected [9]

$$\ldots, \quad (np\sigma_g)^2, \quad (np\pi_u)^4, \quad (np\pi_g)^4, \quad (np\sigma_u)^2,$$

TABLE III. – *Components of the g-tensor, measured at* 9.3 kmc. *Polycrystalline* α, α-*diphenyl* β-*picryl hydrazyl* (DPPH) *served as a standard. The g-factor corresponding to the point of zero slope of the absorption curve was given to us as* 2.0034. *The experimental error in this table does not include a possible error in the g-factor of the reference standard. For* LiF *and* KBr *the uncertainty in the g-factor is mainly due to overlapping lines.*

Crystal	$g_{x'}$	$g_{y'}$	$g_{z'}$	Maximum error
LiF	2.0227	2.0234	2.0031	±0.0010
KCl	2.0428	2.0447	2.0010	±0.0001
NaCl	2.0489	2.0425	2.0010	±0.0001
KBr	2.179	2.175	1.980	± 0.001

In the ground state the hole is in the $np\sigma_u$ level and the configuration of the halogen$_2^-$ molecule-ion is then

$$\ldots, \quad (np\sigma_g)^2, \quad (np\pi_u)^4, \quad (np\pi_g)^4, \quad (np\sigma_u)\,.$$

The symmetry of the crystalline field makes the x' and y' axes non-equivalent (see Fig. 1). Thus the π-levels are slightly split into two levels associated with $p_{x'}$ and $p_{y'}$ functions. The level scheme is shown in Fig. 2. Among the levels shown in this scheme it is only $np\pi_u$ which can give a contribution to the g-shift. Levels outside the ones noted here are probably too far away to affect the g-shift significantly.

Using the perturbation procedure proposed by PRYCE [10] one

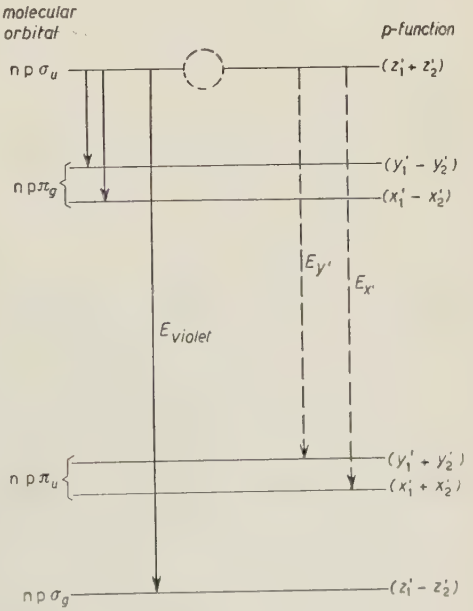


Fig. 2. – Schematic representation of the orbital energy levels of the halogen$_2^-$ molecule-ion. The broken arrows represent the (optically forbidden) transitions that produce the g-shifts, and the full arrows represent the allowed optical transitions.

obtains in second order and neglecting overlap integrals the following expressions for the g-shifts

$$g_{x'} - g_0 \cong \frac{2\lambda\alpha^2}{E_{y'}} ,$$

$$g_{y'} - g_0 \cong \frac{2\lambda\alpha^2}{E_{x'}} ,$$

$$g_z - g_0 \cong 0 .$$

The spin orbit coupling constant λ can be taken from optical data on halide ions, and the fraction α^2 of p-character from Table II. The transition energies are labelled as in Fig. 2. A more elaborate calculation of the g-shifts was made by INUI, HARASAWA and OBATA [11]. These authors took into account the change of the wave function due to the spin orbit interaction which results in the addition of (negative) $1/E^2$ terms in the above expressions for the g-shifts. Using this refined formula and the experimental g-values in Table III one calculates the E values listed in Table IV. The transitions $np\sigma_u \rightarrow np\sigma_g$ and $np\sigma_u \rightarrow np\pi_g$ are optically allowed transitions, the former being in the violet, the latter in the infrared. The violet transition is of particular interest here since V-centers are known to absorb in this spectral region. The calculation of the energy of these transitions is not a simple matter and has not been carried out. However the level scheme depicted in Fig. 2 suggests the following *limits* for E_{violet}

$$E_{y'} \text{ (or } E_{x'}) < E_{\text{violet}} < 2E_{y'} \text{ (or } 2E_{x'}) .$$

TABLE IV. – *Excited states of the halogen$_2^-$ molecule-ion computed from the g-shifts. (Note that the positive sign of the constant « a » is more likely to be the correct choice).*

Crystal	Sign a	$E_{x'}$ 10^4 cm^{-1}	$E_{y'}$ 10^4 cm^{-1}	limits for E_{violet} 10^4 cm^{-1}
LiF	+	1.23	1.28	$1.23 < E_v < 2.56$
	−	1.49	1.54	$1.49 < E_v < 3.08$
KCl	+	1.66	1.74	$1.66 < E_v < 3.48$
	−	2.00	2.08	$2.00 < E_v < 4.16$
NaCl	+	1.74	1.52	$1.52 < E_v < 3.48$
	−	2.12	1.84	$1.84 < E_v < 4.24$
KBr	+	1.73	1.64	$1.64 < E_v < 3.46$
	−	2.24	2.11	$2.11 < E_v < 4.48$

It is interesting to note that among the predominant optical V-bands (V_1, V_2, V_3, V_4) it is only the V_1-band that falls within these limits (*) [4].

5. – Other V-centers.

The V-center described in the preceding sections is the simplest electron deficiency in alkali halides we have encountered so far in the course of our paramagnetic resonance work. The structure determination of two more complicated V-centers is well under way. The first of these is a halogen$_2^-$ molecule-ion associated with a vacancy aggregate (probably a vacancy pair). The second [13] consists of an electron deficiency shared by three halide ions. We know that the three nuclei form a flat, isosceles triangle. The plane of the triangle is a (001)-plane and its base is parallel to a [110]-axis of the crystal. It appears that the structure of this center corresponds to the model which SEITZ proposed for the V_4-center [4].

(*) This contradicts the conclusion reached by M. H. COHEN [12] who erred by a factor of 2 in calculating the energies of the excited states.

REFERENCES

[1] J. H. DEBOER: *Recueil des Trav. Chim. d. Pays-Bas*, **56**, 301 (1937).
[2] A. F. KIP, C. KITTEL, R. A. LEVY and A. M. PORTIS: *Phys. Rev.*, **91**, 1066 (1953); A. M. PORTIS: *Phys. Rev.*, **91**, 1071 (1953).
[3] H. PICK: *Ann. der Phys.*, **31**, 365 (1938).
[4] F. SEITZ: *Rev. Mod. Phys.*, **26**, 7 (1954).
[5] T. G. CASTNER and W. KÄNZIG: *J. Phys. Chem. Solids*, **3**, 178 (1957).
[6] T. O. WOODRUFF and W. KÄNZIG: *J. Physics and Chemistry of Solids* (in press).
[7] D. DUTTON and R. MAURER: *Phys. Rev.*, **90**, 126 (1953).
[8] W. KÄNZIG: *Phys. Rev.*, **99**, 1890 (1955).
[9] G. HERZBERG: *Spectra of Diatomic Molecules*, 2nd ed., p. 329.
[10] M. H. L. PRYCE: *Proc. Phys. Soc.*, A **63**, 25 (1950).
[11] T. INUI, S. HARASAWA and Y. OBATA: *Journ. Phys. Soc. Japan*, **11**, 612 (1956).
[12] M. H. COHEN: *Phys. Rev.*, **101**, 1432 (1956).
[13] M. H. COHEN, W. KÄNZIG and T. O. WOODRUFF: *Phys. Rev.*, **108**, 1096 (1957).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Correlation Effects in Diffusion in Solids.

A. B. LIDIARD

Atomic Energy Research Establishment - Harwell(*)

1. – Introduction.

In recent years there has been an increasing interest in the kinetics, as distinct from the phenomenology of diffusion in crystalline solids. In the phenomenological theory of diffusion one uses Fick's law

$$\boldsymbol{j} = -D\nabla c\,, \tag{1.1}$$

where $\boldsymbol{j}$ is the current density and c the concentration, together with the conservation equation

$$\nabla\cdot\boldsymbol{j} + \frac{\partial c}{\partial t} = 0\,. \tag{1.2}$$

Results are expressed in terms of D which is left as an unknown macroscopic parameter. In the kinetic theory of diffusion one aims to relate D to the atomic parameters which describe the diffusion mechanism, and, conversely, to use experimental information on D to give insight into these mechanisms. One feature of this problem which has recently excited interest is the spatial correlation which may exist between successive movements of a diffusing atom; this is an essentially geometrical feature of the diffusion mechanism and in favorable cases is one which may be isolated from the other factors which make up the observed diffusion coefficient (e.g. atomic jump frequency). Rather direct information about the nature of the mechanism can thereby be obtained.

Let us consider the problem in slightly more detail for the example of an isotope diffusing by a vacancy mechanism. The usual elementary treatments [1] provide an expression for the diffusion constant in an isotropic lattice which

(*) Now at the Department of Physics, University of Reading.

can be written as

$$D_T = \tfrac{1}{6} \Gamma r^2 , \tag{1.3}$$

where r is the distance moved by the atom in each jump (nearest neighbour distance) and Γ is the average number of jumps made by each isotopic atom in unit time. In arriving at this formula it is assumed that the isotopic atoms make a three-dimensional random walk through the crystal: the jump frequency Γ is the product of the probability that there is a vacancy next to the isotope with the frequency that the isotope jumps into such a vacancy. However, although *on the average* the isotopic atoms do jump equally frequently in all directions, when we look at the movements of an isotope over short times, say a sequence on several jumps, we see that successive movements are dependent on the directions of the jumps which have gone before. Take the case where the isotope has just arrived at position P from P'. Of necessity the vacancy is still neighbour to the isotope so that although the *vacancy* makes random jumps, the next jump of the *isotope* is most likely to be back to P' and least likely to be another jump in the forward direction. This correlation between the directions of successive jumps of an atom has been omitted in arriving at equation (1.3). Its inclusion leads to another factor, which we call the correlation factor, f, in (1.3),

$$D_T = \tfrac{1}{6} \Gamma r^2 f . \tag{1.4}$$

We note that this factor f will only appear when the diffusing atom is moved by the intervention of another « particle », in the above example the vacancy. It does not appear in the diffusion coefficient of interstitial solutes, e.g. C in iron, where successive jumps are strictly random and independent of one another.

In this paper we review the calculations of f which have been made for different diffusion mechanisms. The recent interest in these calculations has arisen largely because in ionic conducting solids the Γ factor in D_T can effectively be determined independently (from conductivity); reliable experimental values of f have thereby become available. The steps in our development are:

1) a random walk theory is used to obtain an expression for D_T (equation (1.4)); the jump frequency Γ involves the product of the defect concentration and jump frequency (there may also be a numerical factor);

2) a random walk theory also provides expressions for the diffusion coefficients of the defects themselves $D_{i,v}$, involving their jump frequency but not their concentration;

3) these diffusion coefficients can be related to the corresponding contributions to the ionic conductivity $\sigma_{i,v}$ through the Einstein relation

$$\frac{\sigma_{i,v}}{D_{i,v}} = \frac{n_{i,v}e^2}{kT}\,, \tag{1.5}$$

where n is the number of defects per unit volume and e their charge;

4) elimination of the defect jump frequency and defect concentration between the three equations so obtained yields a value for the quantity $kT\sigma/Ne^2D_T$ ($N =$ number of normal lattice ions per unit volume). This quantity is characteristic of the defect mechanism assumed and may be compared directly with the observed value. It is not equal to unity as a naive application of the Einstein relation would lead one to suppose. This is essentially because the Einstein relation applies only to a dilute solution of charged particles whose distribution in an external potential follows the exponential Boltzmann law. We can therefore apply this relation to the defects themselves (1.5) but not to the normal ions.

Calculations corresponding to step 1 are reviewed in Sect. **2**, **3** and **4**. Steps 2, 3 and 4 which allow one to make contact with experimental results are discussed in Sect. 5. Sect. **6** contains experimental results so far obtained and Sect. **7** considers some possibilities for future work on ionic conductors. The material of this account is largely drawn from references [2-7].

2. – Diffusion coefficients, general.

Firstly we shall consider the approach leading to equation (1.4). In the random walk theory of diffusion one begins by supposing that an atom is initially at the origin of co-ordinates and that after a time t it has made n jumps $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, ..., $\boldsymbol{r}_n$, the total displacement being $\boldsymbol{R}(t) = \sum \boldsymbol{r}_i$. If a large number of atoms were initially at the origin a distribution of displacement vectors $\boldsymbol{R}(t)$ would be obtained. Such a distribution could also be described by the macroscopic laws of diffusion, (1.1), and (1.2), so that, knowing the jump frequency of the atoms and the probability of jumping in the different directions, a comparison of the two descriptions can be made. The macroscopic description gives the solution for this case as

$$c(R, t)\,\mathrm{d}\boldsymbol{R} = \frac{1}{(2\sqrt{\pi Dt})^3} \exp\left[-\frac{R^2}{4Dt}\right]\mathrm{d}\boldsymbol{R}\,, \tag{2.1}$$

for the fraction of atoms in the volume $\mathrm{d}\boldsymbol{R}$ at $\boldsymbol{R}$. The second moment of this

distribution is

$$R^2(t) = 6Dt \,. \tag{2.2}$$

We will now compare (2.2) with the mean square displacement calculated directly and so derive an equation for D; but we will not consider the further question of the consistency of the higher moments, i.e. we shall accept Fick's law (1.1) as always valid for isotropic diffusion in an ideal system. We write

$$\begin{aligned} R^2(t) &= (\sum \boldsymbol{r}_i)^2 \\ &= \sum r_i^2 + 2 \sum_{j=1}^{n-1} \sum_{i=1}^{n-j} \overline{\boldsymbol{r}_i \cdot \boldsymbol{r}_{i+j}} \,. \end{aligned} \tag{2.3}$$

For situations where all jump vectors are of equal length the first term in (2.3) is simply nr^2. If this term alone were present then (2.2) and (2.3) give (1.3) when we set $n/t \equiv \Gamma$, the jump frequency. It is the second term which exe presses the correlation effects; its magnitude depends on the mechanism w- consider and the crystal lattice in which diffusion is taking place.

3. – Diffusion coefficients for isotopes and foreign atoms, vacancy mechanism.

When a vacancy mechanism is operative in a lattice where all jumps are equivalent (e.g. cubic lattices) the average value of $\boldsymbol{r}_i \cdot \boldsymbol{r}_{i+j}$ is clearly independent of i. Equation (2.3) can then be written as

$$R^2(t) = nr^2 + 2r^2 \sum_{j=1}^{n-1} (n - j) \cos \overline{\theta_j} \qquad \text{(vacancy)} \,. \tag{3.1}$$

Further, it is not difficult to show (5) that when there is two-fold or three-fold rotational symmetry about the jump vectors then

$$\overline{\cos \theta_j} = (\overline{\cos \theta_1})^j \qquad \text{(vacancy)} \,, \tag{3.2}$$

so that for large n (3.1) becomes

$$\overline{R^2(t)} = nr^2 \frac{(1 + \cos \theta_1)}{(1 - \cos \theta_1)} \qquad \text{(vacancy)} \,, \tag{3.3}$$

By (2.2) therefore

$$D = \frac{1}{6} \Gamma r^2 \frac{(1 + \overline{\cos \theta_1})}{(1 - \overline{\cos \theta_1})} \qquad \text{(vacancy)} \,. \tag{3.4}$$

The computation of the correlation factor in these cases is therefore reduced to the determination of the average $\cos\theta$ between one atom jump and the next. One carries out this evaluation by following the jumps of the *vacancy*, noting the probabilities of return to the atom from the different directions θ allowed by the lattice. Analytical methods and methods based on the analogy with electrical networks (electrical currents taking the place of vacancy currents) have been elegantly exploited by COMPAAN and HAVEN who have derived almost all the recent results in this field. Some of their results are given in Table I.

TABLE I. – *Correlation factors for isotopic diffusion via free vacancies* (after COMPAAN and HAVEN [5]).

	$\overline{-\cos\theta_1}$	$f = \dfrac{1+\overline{\cos\theta_1}}{1-\overline{\cos\theta_1}}$
2 *dimensional*		
honeycomb layer	$\frac{1}{2}$	$\frac{1}{3}$
square layer	0.363 28	0.467 05
hexagonal layer	0.282 00	0.560 06
3 *dimensional*		
diamond lattice	$\frac{1}{3}$	$\frac{1}{2}$
simple cubic lattice	0.208 10	0.655 49
body-centred cubic lattice	0.161 78	0.721 49
face-centred cubic lattice	0.122 68	0.781 45

It may be noted [2] that in a linear chain there is no diffusion; $\overline{\cos\theta_j}$ is necessarily $(-1)^j$ and it is seen that the correlation term just cancels the leading term nr^2 (n even). This is an example where correlations are overwhelmingly important.

Another example where correlation effects may make an important change in the diffusion coefficient is provided by the diffusion of foreign atoms which bind vacancies to themselves, e.g. in the alkali halides divalent ions may strongly attract vacancies for essentially electrostatic reasons [8, 9]. If we suppose that we have a face-centred cubic lattice (which could be one sublattice in a NaCl or zinc blende structure) and that the distance between nearest neighbours is $\sqrt{2}\,a$ then the impurity diffusion constant is found [5, 6, 10] to be

$$D = \frac{a^2}{3}\,\frac{w_1 w_2}{w_1 + w_2}\,, \tag{3.5}$$

where w_2 is the jump frequency for impurity-vacancy exchange and w_1 is the jump frequency for an attached vacancy going from one position to another. The correlation factor is $w_1/(w_1+w_2)$ and shows that unless the vacancy can circle round the impurity then the impurity atom just jumps back and forth with no net displacement, similarly to the one-dimensional example.

The symmetry required for equation (3.2) to follow is present in the examples discussed above, i.e. isotopic diffusion by free vacancies or impurity diffusion by bound vacancies. An example in which it is not present is provided by isotopic diffusion caused by vacancies bound to impurity atoms. The presence of the impurity atom destroys the rotational symmetry about the isotope-vacancy axis which would have existed otherwise. In such cases one must work with the full equation (3.1) and the calculations become correspondingly more involved. COMPAAN and HAVEN [4] were however able to obtain unmerical values for the correlation factor by using appropriate electrical analogues. Their results are shown in Fig. 1. It would be interesting to know the simple limiting forms of these correlation factors for small w_2/w_1 and w_-/w_+.

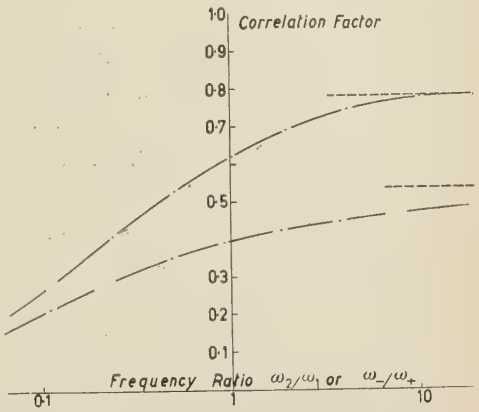


Fig. 1. – Correlation factors for isotope diffusion in a NaCl lattice caused by impurity-vacancy pairs (lower curve) and cation vacancy-anion vacancy pairs (upper curve). The jump frequencies of the bound cation and anion vacancies are w_+ and w_- respectively. (After COMPAAN and HAVEN [5]).

4. – Diffusion coefficients of isotopes, interstitial mechanism.

It is known from electrical conductivity studies that AgBr and AgCl display predominantly cationic Frenkel disorder [11, 12] and we therefore consider the correlation effects which may appear in connection with interstitial mechanisms. Firstly as we have remarked above, interstitial solid solutions show no effects, the solute atoms jumping randomly by themselves from one site to another. Likewise, diffusion of Ag^+ isotopes in AgBr will take place without correlation effects if the interstitials move directly and randomly from one interstitial site to another.

However it has always appeared possible and now seems confirmed [3, 7] that the Ag^+ interstitials in AgBr and AgCl move by the so-called « intersti-

tialcy » or « indirect-interstitial » mechanism. In this the interstitial ion moves to one of the surrounding normal sites and pushes the normal ion there into the next interstitial position. Clearly if the isotope is present as the first interstitial its next jump, after the move to the normal site, is going to be quite strongly correlated with the first. For, the first move of the newly created interstitial may be straight back to its previous position (probability $\frac{1}{4}$ in the AgBr lattice). On the other hand when the isotope is present initially as the normal ion which is displaced into the interstitial state then its next jump is *un*correlated with the first. In this example then only alternate pairs of jumps are correlated (i.e. pairs of the type interstitial $\rightarrow$ normal $\rightarrow$ interstitial). Hence of the terms $\boldsymbol{r}_i \cdot \boldsymbol{r}_{i+j}$ in equation (2.3) all those with $j > 1$ are zero and of those for which $j = 1$ only alternate ones are non-zero. Hence

$$\overline{R^2(t)} = nr^2(1 + \overline{\cos\theta}) , \qquad \text{(indirect-interstitial)} \tag{4.1}$$

so that

$$D = \tfrac{1}{6}\Gamma r^2(1 + \overline{\cos\theta}) , \tag{4.2}$$

where $\cos\theta$ is the average cosine of the angle between consecutive interstitial $\rightarrow$ normal and normal $\rightarrow$ interstitial jumps, and as before is a negative quantity.

For the NaCl type of lattice (AgBr, AgCl) $\cos\theta$ has [13] the value $-\frac{1}{3}$ if we consider only acts in which the two atoms move in one line e.g. $(\frac{1}{2}\,\frac{1}{2}\,\frac{1}{2}) \rightarrow$ $\rightarrow (111) \rightarrow (\frac{3}{2}\,\frac{3}{2}\,\frac{3}{2})$. However inspection of the NaCl lattice shows that two other possibilities exist, namely acts of type $(\frac{1}{2}\,\frac{1}{2}\,\frac{1}{2}) \rightarrow (111) \rightarrow (\frac{1}{2}\,\frac{3}{2}\,\frac{3}{2})$ and of type $(\frac{1}{2}\,\frac{1}{2}\,\frac{1}{2}) \rightarrow (111) \rightarrow (\frac{1}{2}\,\frac{1}{2}\,\frac{3}{2})$. It has been shown [14] that if only the first of these two types occurs then $\cos\theta$ is $-0.030\,303$ whilst if only the second occurs then $\cos\theta = -0.036\,17$. The effect is smaller for the non-collinear jumps because for a given movement of the first ion there are three equivalent movements for the second, whereas for the collinear jump there is only one movement.

5. – Electrical conductivity and its relation to the tracer diffusion coefficient.

By following the random walk of a vacancy or of an interstitialcy we can derive diffusion coefficients for the defects themselves. Equations (2.2) and (2.3) can be employed and, since there are no correlations to be considered, we get simply

$$D_{i,v} = \tfrac{1}{6}\Gamma_{i,v} r_{i,v}^2 . \tag{5.1}$$

Hence by the Einstein relation the corresponding contributions to the ionic conductivity are

$$\sigma_{i,v} = \Gamma_{i,v} n_{i,v} r_{i,v}^2 e^2/6kT \,. \tag{5.2}$$

By adding up these contributions we obtain the total conductivity which may then be compared with the total tracer diffusion coefficient calculated from equations (3.4) and (4.2).

Let us make this comparison for the Frenkel model applicable to AgCl and AgBr. We assume first that the collinear interstitialcy mechanism operates. Then Γ for the tracer ions (equation (4.2) is $2n_i\Gamma_i/N$, where N is the number of normal ions per unit volume. The factor 2 appears since two ions move in each interstitialcy jump and either may be a tracer. If a is the anion-cation separation the jump distance of the tracer is $\sqrt{3}\,a/2$. Hence, substituting the value $\frac{2}{3}$ of the correlation factor (4.2) gives

$$D_{Ti} = \frac{1}{6}\frac{n_i}{N}\Gamma_i a^2 \,. \tag{5.3}$$

The jump distance of the interstitialcy is $\sqrt{3}a$ so that (5.2) yields

$$\sigma_i = \frac{\Gamma_i n_i e^2 a^2}{2kT} \,,$$

so that

$$\frac{\sigma_i}{D_{Ti}} = 3\,\frac{Ne^2}{kT} \,. \tag{5.4}$$

The coefficient 3 is made up of the correlation factor $\frac{2}{3}$ in the denominator and of (opposed) terms coming the fact that the interstitial jump distance is twice the tracer jump distance and that the tracer jump frequency is twice $n_i\Gamma_i/N$.

More generally when all three types of interstitialcy jump are operative—in proportions p_1 for collinear, p_2 when the two displacements are at $\cos^{-1}(\frac{1}{3})$ and p_3 when they are at $\cos^{-1}(-\frac{1}{3})$—we have

$$\frac{\sigma_i}{D_{Ti}} = \frac{Ne^2}{f_i kT} \,, \tag{5.5}$$

where

$$f_i = \frac{(1+\cos\theta)}{2(p_1 + 2p_2/3 + p_3/3)} \,, \qquad (p_1+p_2+p_3=1). \tag{5.6}$$

Friauf [7] has evaluated $\cos\theta$ approximately as a function of p_2/p_1 when $p_3 = 0$. More extensive calculations have recently been carried out by Compaan and Haven for the complete range of p_1, p_2, p_3 values.

Before closing this section we should note that a rather different model for the Frenkel disorder in AgCl has been put forward by HOVE [15]. It was suggested that the stable interstitial configuration might be a sort of dumb-bell arrangement of two ions about a normal lattice point. These dumb-bells would have (111) orientation, and movement would occur by a jump of one of the ions of the pair to another normal site of the lattice cell where the new dumb-bell is formed. It was shown independently by MCCOMBIE and by COMPAAN and HAVEN that this model leads to the same equation (5.4) as does the collinear jump model for the conventional interstitial.

The observed values of σ and D_T will include contributions from the vacancies as well as the interstitials,

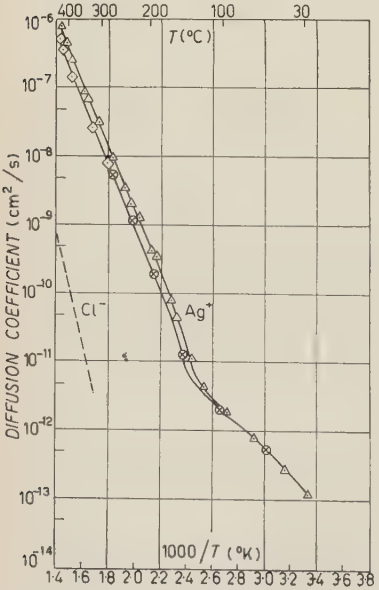
$$\frac{\sigma_v}{D_{Tv}} = \frac{Ne^2}{f_v kT} . \tag{5.7}$$

Hence

$$\frac{\sigma}{D_T} = \frac{\sigma_i + \sigma_v}{D_{iT} + D_{Tv}} = \frac{Ne^2}{kT} \frac{1}{(t_i f_i + t_v f_v)} ,$$

where t_i and t_v are the interstitial and vacancy transport numbers.

6. – Experimental results for AgBr and AgCl.



The results of COMPTON and MAURER [3] obtained on AgCl show quite unambiguously that the ratio $kT\sigma/Ne^2D_T$ is greater than unity in the region of intrinsic conduction (Fig. 2). The ratio is about 2 at 200 °C decreasing to about 1.7 at 350 °C. The fractions of current carried by the interstitials and vacancies, t_i and t_v, can easily be obtained from the mobility ratios

Fig. 2. – Diffusion coefficients in AgCl plotted against the reciprocal of the absolute temperature. The upper full line drawn through the points △ is $kT\sigma/Ne^2$ where σ is the measured ionic conductivity. The lower full line drawn through the points ◈ and ⊗ is D_T for ^{110}Ag. Both sets of measurements were made on a specimen grown and annealed in air. The D_T values are in good agreement with those measured in crystals grown under different conditions, and the intrinsic conductivities are in good agreement with earlier results reported by EBERT and TELTOW [12]. The dashed curve is the diffusion coefficient of ^{36}Cl as measured in three different crystals.

determined by EBERT and TELTOW [12]. Also f_v is known to be 0.78 for the face-centred cubic cation sub-lattice (Table I). In this way COMPTON and MAURER find the observed ratios $kT\sigma/Ne^2D_v$ to be consistent with $f_i = 0.5$. This is somewhat larger than $\frac{1}{3}$ and indicates the occurrence of non-collinear interstitialcy jumps. Or alternatively if one is thinking in terms of Hove's dumb-bell model then it is necessary to suppose that the dumb-bells can undergo hindered rotations from one (111)-orientation to another.

Very similar results have been obtained for AgBr by FRIAUF [7]. Using mobility ratios found by TELTOW [11], FRIAUF infers f_i values from his data and then, from calculations of f_i as a function of p_1/p_2 (p_3 assumed zero), determines p_1/p_2 at various temperatures. The numbers so obtained can be fitted to the formula

$$\frac{p_2}{p_1} = 38 \exp\left[-\frac{0.15}{kT}\right] \text{eV} .$$

Knowing the total interstitial mobility [11], the contributions of the collinear and the non-collinear processes can then be separated. In this way FRIAUF finds an activation energy of 0.08 eV for the collinear jumps and of 0.23 eV for the non-collinear jumps.

7. – Possibility of observing effects in other systems.

Substances with the fluorite (CaF_2) lattice display predominantly anionic Frenkel disorder. Interstitialcy movements in such a lattice can occur in only one way, the two atomic displacements making an angle $\cos^{-1}(\frac{1}{3})$ with each other. However the correlation factor $(1 + \cos\theta)$ is only slightly different from unity, so $kT\sigma_i/Ne^2D_{Ti}$, which is 1.4, is made up almost entirely from the factor $\frac{4}{3}$ resulting from the displacement-distance and two-ions-in-a-jump effects. The ratio $kT\sigma_v/Ne^2D_{Tv}$ for vacancies is 1.53 (the anion sub-lattice is simple cubic, Table I) so that there is little difference here between the two mechanisms. Interstitials can be made predominant in these systems by doping with a cation of higher valency. Unfortunately there appears to be only one compound which is suitable for tracer studies, $SrCl_2$.

Somewhat more striking effects may be observable in ionic conductors with the CsCl structure (e.g. TlCl, TlBr, TlI, CsCl, CsBr, CsI). Most of what follows is speculative since these systems have not yet been studied very completely, and the type of disorder and even the conductivity mechanism may be in doubt [16, 17]. If they show Schottky disorder then $kT\sigma/Ne^2D_T$ will be 1.53, as appropriate to a vacancy mechanism in a simple cubic lattice. However Frenkel disorder, which is rather probable since it is favoured by the large Van der Waals attraction associated with the heavy ions Tl^+ and

Cs^+, presents more interesting possibilities. Since the anions (Cl^-, Br^- and I^-) are larger than the cations (Tl^+ and Cs^+) one would expect cationic Frenkel disorder and one would expect the stable interstitial cation sites to be those lying at the centre of a face of the elementary anion cube, i.e. midway between two cations on a (100) axis. Collinear interstitialcy jumps in the (100) directions may then occur relatively easily (frequency ν_1). However if only these jumps occur there can be no tracer diffusion, since we have then just another example of one-dimensional diffusion. Non-collinear jumps, which here occur at right angles (e.g. $(0\,\frac{1}{2}\,\frac{1}{2}) \rightarrow (\frac{1}{2}\,\frac{1}{2}\,\frac{1}{2}) \rightarrow (\frac{1}{2}\,1\,\frac{1}{2})$), must take place if there is to be any tracer diffusion. If the frequency of these non-collinear jumps (ν_2) is low relative to ν_1 then D_{Ti} will be proportional to ν_2, so that $kT\sigma_i/Ne^2D_{Ti}$ will be proportional to ν_1/ν_2 and may therefore be very large. On the other hand, at high values of ν_2, $\cos\theta$ will be very small, since for ν_1 zero $\cos\theta$ is zero also—a result which follows quite simply from the symmetry imposed by the perpendicular ν_2 jumps. In this limit $kT\sigma_i/Ne^2D_{Ti}$ is unity, the factor 2 coming from the fact that the insterstitialcy displacement length is $\sqrt{2}$ times the tracer displacement length just cancelling with the factor coming from the fact that two atoms jump in each act ($\Gamma_T = 2n_i\Gamma_i/N$). However the limit $\nu_1 \ll \nu_2$ seems rather unlikely to occur and we may therefore look forward to the results of a thorough study of the CsCl systems in the hope of seeing evidence of the rather strong correlation effects which the interstitialcy mechanism gives rise to here.

8. – Summary.

Here we list the main points which have so far emerged:

1) Correlation effects occur when the diffusing species (tracer, impurity) is moved by another particle (vacancy, interstitial).

2) Correlation effects in linear systems are so large that no diffusion occurs.

3) Correlation effects may also be important in impurity diffusion.

4) Correlation effects with tracer diffusion in ionic conductors can be seen in a comparison of $kT\sigma/Ne^2$ with D_T. For some interstitial mechanisms the ratio $kT\sigma/Ne^2D_T$ is greater than unity not only because of the correlation effect but also because the interstitialcy jump distance differs from the tracer jump distance.

5) « Indirect interstitial » or « interstitialcy » migration occurs in AgCl and AgBr, but it is necessary to suppose that non-collinear jumps occur as well as collinear jumps.

6) Hove's dumb-bell model without rotations leads to the same ratio for $kT\sigma_i/Ne^2D_{Ti}$ as does the collinear interstitialcy mechanism.

7) Correlation effects in CsCl lattices may be very large if Frenkel disorder occurs.

* * *

In conclusion I like to record my gratitude to the Italian Physical Society for their generous support during the period of the 1957 Summer School.

REFERENCES

[1] See e.g. C. ZENER: *Imperfections in Nearly Perfect Crystals* (New York, 1952), p. 289, or A. H. COTTRELL: *Theoretical Structural Metallurgy*, 2nd Ed. (London, 1955).

[2] J. BARDEEN and C. HERRING: *Imperfections in Nearly Perfect Crystals* (New York, 1952), p. 261. This paper was the first in which the correlation effect was pointed out and discussed theoretically.

[3] W. D. COMPTON: *Phys. Rev.*, **101**, 1209 (1956) and W. D. COMPTON and R. J. MAURER: *J. Phys. Chem. Solids*, **1**, 191 (1956).

[4] C. W. MCCOMBIE and A. B. LIDIARD: *Phys. Rev.*, **101**, 1210 (1956).

[5] K. COMPAAN and Y. HAVEN: *Trans. Faraday Soc.*, **52**, 786 (1956). Also *Faraday Society Discussion*, **23**, 105 (1957) and *Proceedings of the 3rd International Conference on the Reactivity of Solids held in Madrid* in April 1956, p. 255.

[6] A. D. LE CLAIRE and A. B. LIDIARD: *Phil. Mag.*, (8) **1**, 518 (1956).

[7] R. J. FRIAUF: *Phys. Rev.*, **105**, 843 (1957).

[8] F. BASSANI and F. G. FUMI: *Nuovo Cimento*, **11**, 274 (1954); F. G. FUMI and M. TOSI: *Faraday Society Discussion*, **23**, 92 (1957).

[9] M. CHEMLA: *Ann. Phys. Paris*, (13), **1**, 959 (1956), especially chapters V, VI and VII.

[10] A. B. LIDIARD: *Phil. Mag.*, **46**, 815 and 1218 (1955), where it is shown that a concentration-dependent term must be added to (3.5) if the concentration of impurity-vacancy pairs is different from the total impurity concentration.

[11] J. TELTOW: *Ann. Phys.*, **5**, 63 and 71 (1949).

[12] I. EBERT and J. TELTOW: *Ann. Phys.*, **15**, 268 (1955).

[13] The value -0.33 was obtained first by C. W. MCCOMBIE. Later COMPAAN showed it was exactly $-\frac{1}{3}$.

[14] K. COMPAAN and Y. HAVEN: Private communication.

[15] J. E. HOVE: *Phys. Rev.*, **102**, 915 (1956). We believe HOVE's calculation of the activation energy of the non-collinear process to be too high since it appears to take the interaction between the two ions as e^2/r, whereas $e^2/\varepsilon r$ would appear more appropriate.

[16] W. W. HARPUR, R. L. MOSS and A. R. UBBELHODE: *Proc. Roy. Soc.*, A **232**, 196 (1955); W. W. HARPUR and A. R. UBBELHODE: *Proc. Roy. Soc.*, A **232**, 310 (1955).

[17] K. HAUFFE and A.-L. GRIESSBACH-VIERK: *Zeits. f. Elektrochemie*, **57**, 248 (1953).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

On Some Physical Consequences of the Non-Linear Theory of Dislocations.

A. SEEGER

Max-Planck-Institut für Metallforschung - Stuttgart

1. – Introduction and general discussion.

In this lecture we shall not take up a new subject but rather carry further the discussion of a number of topics that have been touched in other lectures; in particular in those on «Dislocation Theory» by Professor FRANK. The common feature of these topics will be that they cannot be treated adequately by the application of the linear theory of elasticity to dislocations in crystals. Professor FRANK has already discussed the insufficiency of the linear theory of elasticity to describe the core of a dislocation line satisfactorily. Most of the phenomena that we will discuss are usually associated with the dislocation core, i.e. the very heavily distorted region along the center of the dislocation line, the diameter of which is of the order of magnitude of the dislocation strength b. We will see, however, that it does not suffice to restrict our attention to the core. In fact, the main contributions to some of the phenomena we shall treat (e.g. the influence of the dislocations on the mean density of the crystals) does not come from the dislocation core itself but from the *environment* of the dislocation core.

Professor FRANK has mentioned various attempts to treat the dislocation core in an improved way, in particular the model first used by PEIERLS and later extended by a number of authors, and the paper by HUNTINGTON, DICKEY and THOMSON [1] on dislocations in ionic crystals like NaCl. These treatments, however, in their present form assume that outside the dislocation core linear theory of elasticity (based on infinitesimal strains and the validity of Hooke's law) holds. They are therefore subject to a general theorem derived by COLONNETTI [2], which states that within the framework of linear elasticity

a state of self-stress leaves the volume and therefore the mean density of an elastic body unchanged. A strained elastic body is said to be in a state of self-stress or self-strain if no external forces are acting on it. The elastic model of an edge or screw dislocation forms an example: the dislocation is generated by cutting a hollow cylinder along a plane containing the axis, shifting the two sides of the cut relative to one another by the Burgers vector, welding them together and removing all external forces that had to be applied in order to deform the cut cylinder. The validity of Colonnetti's law is easily verified for dislocations. According to the linear theory of elasticity the stress field of a *screw* dislocation contains no dilatation at all, whereas the dilatation and the compression on opposite sides of the glide plane of an *edge* dislocation are the same in magnitude and cancel each other exactly when the total change in volume is determined by an integration over the whole crystal. On account of Colonnetti's theorem none of the treatments so far mentioned is therefore able to give the change in density that comes from the environment of the dislocation core.

In principle the effect we are looking for could be obtained by treating not only the dislocation core but the whole crystal by an atomistic theory. From a practical point of view, however, this is rather undesirable, since such a treatment would be prohibitively complicated and would furthermore involve functions like the interaction potentials between atoms, which in most cases are not sufficiently well known.

We are more interested in a simpler treatment which is based on continuum theory and involves parameters that can be determined experimentally. A treatment of this kind, which does allow for the effects that dislocations have on the crystalline density and related properties, can be based on the use of the non-linear theory of elasticity. The relation between the change in density due to self-stresses and the anharmonicity of the stress-strain relations has been known for a long time (see e.g. a paper by ZENER [3] and its discussion by SEITZ); it has not been applied to dislocation theory, however, until very recently.

The basic idea is demonstrated in Fig. 1 showing the potential energy of a crystal as a function of the interatomic distance d. In the environment of the equilibrium position the $U(d)$ curve can be well represented by a parabola; in this region Hooke's law holds; the non-linear effects will be small. If we consider large compressions or dilatations, however, we get deviations from Hooke's law. They are always

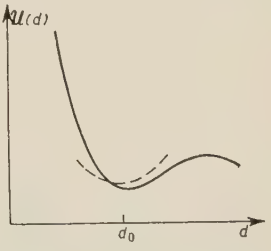


Fig. 1. The potential energy per atom U of a crystal as a function of the distance between atoms, d_0 = equilibirum distance. The dotted parabola corresponds to a harmonic law of force.

in such a direction that the crystal becomes harder under increasing compression, and softer under increasing dilatation. This means that the symmetry between dilatation and compression is upset, and that the deviations from Colonnetti's theorem which we expect when allowing for the higher order effects are such as to give a *decrease* in the average density of the crystal.

Before we proceed to a more detailed discussion of the density change in Sect. 2, we should like to mention that these effects are important not only in themselves, but that they have an important consequence for the charge distribution around a dislocation line and thereby for the electrical properties of dislocations, e.g. the electrical resistivity due to dislocations in metals.

For the sake of simplicity let us consider the conduction electrons of a metal as being a gas of quasi-free electrons with an effective mass m^*, the Fermi energy of the gas being F. For the mean energy $\bar{\varepsilon}$ of the electrons we obtain

$$\bar{\varepsilon} = \frac{3}{5} F = \frac{3}{10} \frac{\hbar^2}{m^*} k_F^2 , \tag{1}$$

where $\hbar$ is Planck's constant divided by 2π and k_F is the wave number at the Fermi surface. For a given number of electrons per atom k_F is inversely proportional to the interatomic distance d or the lattice constant. Therefore the mean energy of the electrons is proportional to the $(-\frac{2}{3})$ power of the specific volume V, or

$$\bar{\varepsilon} \simeq \frac{1}{d^2} \simeq V^{-\frac{2}{3}} . \tag{2}$$

This means that the electrons tend to flow into regions of dilatation, charging them negatively with respect to the rest of the metal. Likewise they tend to avoid compressed regions, leaving them positively charged. Of course this unbalance of charge will be screened by the electrons themselves; the problem is therefore one of self-consistency. Combining these considerations with the linear theory of elasticity it has been predicted that a screw dislocation is not electrically charged at all, whereas an edge dislocation should be equivalent to an electrical line dipole due to regions of compression and dilatation above and below the glide-plane [4].

Allowing for the density effects discussed above we see now that both types of dislocations will actually behave like negatively charged line poles. This obviously has a strong effect on the scattering of electrons and on the electrical interaction between dislocation lines and foreign atoms. We shall briefly report on some of these problems in Sect. 3.

The phenomena we have mentioned so far were outside the realm of the linear theory of elasticity but could still be described by a continuum approach, e.g. second order elasticity. There are however interesting dislocation problems

which cannot be described adequately on such a basis. The best known of these problems, in which the atomic structure has to be taken into account explicitly, is the problem of the Peierls stress or, as it is sometimes called, the Peierls-Nabarro force. It has been touched upon in Professor FRANK's lectures. We shall not treat here the *theory* of the Peierls stress which has been carried further by DIETZE [5] but does not yet allow a calculation of the magnitude of the Peierls stress from first principles. Instead we shall briefly report on a recent determination of the Peierls stress in some f.c.c. metals from internal friction experiments (Sect. 4).

2. – The density change due to dislocations.

The strain field of a screw dislocation in an isotropic medium, according to the linear theory of elasticity, corresponds to pure shear and is therefore simpler than that of an edge dislocation which involves also dilatations and compressions. For the sake of simplicity we shall consider screw dislocations in more detail than edge dislocations.

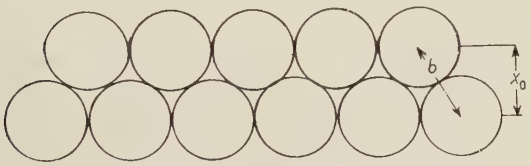


Fig. 2. – Two close-packed rows of atoms in a (111) plane of a f.c.c. metal crystal.

Fig. 2 shows what happens if a close-packed plane is heavily sheared as is the case near the center of a screw dislocation. Due to the fact that there is a strong repulsive potential between neighbouring ion cores, under a shear strain γ neighbouring close-packed rows of atoms will have a separation x which is larger than the equilibrium distance x_0. A dilatation is therefore connected with large « shears » of the type shown in Fig. 3. As an approximation we may put the dilatation proportional to the square of the shear strain, since the effect is the same if the direction of shear is reversed. We can write

$$\Delta = \varkappa\gamma^2 \,. \tag{3}$$

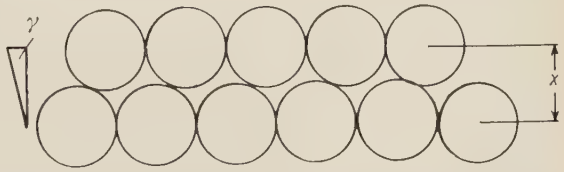


Fig. 3. – Same as Fig. 2, however heavily sheared. The distance x between neighbouring rows is larger than x_0.

Eq. (3) shows that we are dealing with an effect not contained in the linear theory, as discussed in Sect. 1. The constant $\varkappa$ has been determined for a specific model by STHELE and SEEGER [6]; they found

$$\varkappa = \left(\frac{\alpha^2}{12(\alpha+1)} - \frac{1}{4}\right). \tag{4}$$

α is a parameter which determines how rapidly the repulsive interaction between two ion cores varies as a function of their separation in a repulsive potential of the type

$$V_I = A \exp\left|-\alpha \frac{r}{b}\right| . \tag{5}$$

In eq. (5) A is a constant the magnitude of which is unimportant for our present purposes, r is the distance between the ion cores and b is their equilibrium separation (which in a f.c.c. lattice is equal to the dislocation strength). In a metal like copper α will be about 15; therefore $\varkappa$ should be of the order of 1.

The preceding discussion is superior to the purely phenomenological approach which we will describe later in that it allows to estimate an upper limit for the maximum dilatation possible. Even for very large shear strains it is obvious that two neighbouring rows of atoms cannot be separated further than getting « on top of each other » (Fig. 4). We shall make use of this in equation (7).

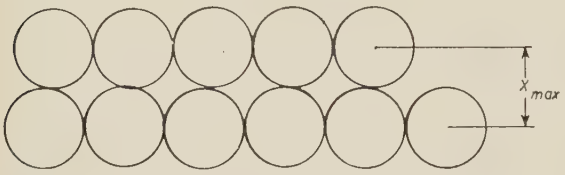


Fig. 4. – Same as Fig. 3; the shear strain corresponds to the maximum separation x_{max} which is possible.

If cylindrical co-ordinates (z, ϱ, φ) are adopted a screw dislocation in an isotropic medium has only one non-vanishing essential shear component

$$\gamma = \gamma_{z\varphi} = \frac{b}{2\pi} \frac{1}{\varrho} , \tag{6}$$

Here ϱ denotes the distance from the dislocation line. Inserting eq. (6) (which was obtained from linear elasticity) in eq. (3) we obtain for the dilatation

$$\Delta = \begin{cases} \dfrac{\varkappa b^2}{4\pi^2} \dfrac{1}{\varrho^2} & \varrho > \varrho_0 \\ \dfrac{\varkappa b^2}{4\pi^2} \dfrac{1}{\varrho_0^2} & \varrho \leqslant \varrho_0 . \end{cases} \tag{7}$$

In eq. (7) we have « cut off » the dilatation in the dislocation core by postulating that it cannot exceed the maximum dilatation discussed above. The cut-off radius is of the order $\varrho_0 \cong b/3$.

The volume change per unit length of the screw dislocation is given by

$$\Delta V = 2\pi \int_0^R \Delta(\varrho) \varrho \, \mathrm{d}\varrho = \frac{\varkappa b^2}{2\pi} \log\left(\frac{R}{\varrho_0} + \frac{1}{2}\right) . \tag{8}$$

In order to avoid the divergence of eq. (8) at the upper limit we had to introduce an outer cut-off radius R. This complication is well known from the calculations of the line energy of a straight dislocation line. This is not accidental; we shall see presently that the change in density and the line energy are closely related to one another. We therefore choose R in the same way as we do in calculating the line energy, i.e. of the order of half the distance between neighbouring dislocations of opposite sign.

Depending somewhat on the dislocation arrangement, R will be of the order 10^{-4} cm to 10^{-5} cm. Since ϱ_0 is about 1 Å and $\varkappa \approx 1$ the volume change $b\Delta V$ of a dislocation line of length b is about one atomic volume or slightly larger. From the point of view of the influence on the crystalline density a screw dislocation behaves similar to a row of vacancies, although the spacial distribution of the dilatation is entirely different in these two cases.

We shall now show how eq. (3) can be extended to those cases in which the linear theory gives more than one essential strain component. Neglecting higher order terms, the average dilatation $\Delta_{\text{ave.}}$ will be a linear function of the squares and the products of the strain components. ZENER [3] has given the following very elegant method to determine this function. $\Delta_{\text{ave.}}$ is a scalar quantity: in an isotropic medium it must therefore be a linear combination of scalar quantities. Out of the six strain components we can form two scalar invariants, e.g. those corresponding to the average energy density $W_{\text{hyd.}}$ due to the hydrostatic stress and the average energy density $W_{\text{sh.}}$ due to the shear stresses. The average dilatation can therefore be written in the form:

$$\Delta_{\text{ave.}} = C_1 W_{\text{hyd.}} + C_2 W_{\text{sh.}} . \tag{9}$$

Here C_1 and C_2 are parameters which contain the elastic constants and, in order to account for the deviations from Hooke's law, the pressure dependence of the elastic constants.

ZENER obtains

$$\begin{cases} C_1 = -\dfrac{1}{K}\left(\dfrac{\mathrm{d}\ln K}{\mathrm{d}\ln V} + 1\right) \\ C_2 = -\dfrac{1}{K}\left(\dfrac{\mathrm{d}\ln G}{\mathrm{d}\ln V} + 1\right). \end{cases} \tag{10}$$

In eq. (10) K and G are the moduli of compression and shear; V denotes the specific volume. If we use the experimental data on the pressure dependence of the elastic constants eq. (9) and (10) give about the same results as the earlier treatment. For extended dislocations we find a slightly smaller change in volume than for unextended dislocations; edge dislocations give a somewhat larger change than screw dislocations.

We mention that eq. (9) can easily be extended to anisotropic media. The right hand side of eq. (9) is found to contain as many terms as there are essentially different elastic constants, e.g. 3 in a cubic crystal.

The preceding results have been compared with experiments [7] in which the change in volume and the stored energy due to dislocation has been measured (*). The agreement is quite good; for details see SEEGER and STEHLE [8] and the Appendix of the present paper.

Since the change in density due to dislocations is distributed over a rather large volume it should give rise to X-ray effects, namely an increase of the average lattice parameter as measured by X-rays, and to a small angle scattering of X-rays and neutrons. The first effect is difficult to measure since the presence of the dislocations not only *shifts* Debye-Scherrer rings and Laue spots but also *broadens* them. A decrease of the lattice parameter of polycrystalline copper during recrystallization (i.e. removal of dislocations) of the right order of magnitude has been found by CRUSSARD and AUBERTIN [9]. We think that this is at least partly due to the effect we have discussed.

The theory of the small angle scattering due to the non-linear dilatations has been worked out both for unextended and for extended screw-dislocations [10]. The predicted effects are difficult to observe with X-rays, since there are other causes of small angle scattering, in particular the double Bragg reflections studied by BEEMAN, NEYNHABER, BRAMMER and WEBB [11]. Using neutrons of sufficiently long wavelengths the double scattering mechanism can be avoided. Under these conditions ATKINSON, working at Harwell, has found small angle scattering of cold-worked metals of the right order of magnitude to be explained by dislocations [12]. No detailed comparison with dislocation theory has yet been made.

3. – Electrical resistivity of dislocations.

As mentioned in Sect. 1 the non-linear dilatation around a dislocation line has a big effect on the electrical resistivity of dislocations. Since the subject has been dealt with elsewhere in some detail [8, 13] we shall confine ourselves to a few remarks. Table I gives the resistivity of N parallel screw dislocation lines per unit area according to various treatments. HUNTER and NABARRO [14] estimated the resistivity due to the shear strains around a screw dislocation by using Born's approximation; SEEGER and STEHLE [8] treated the same problem exactly and obtained a somewhat smaller value. Including the non-linear dilatation and thereby most of the core effects increases this estimate

(*) This type of comparison eliminates the uncertainties connected with the cut-off radius R.

TABLE I. – *Resistivity $\Delta\varrho$ of screw dislocations in copper (not allowing for stacking faults, however) according to various treatments. N = number of dislocation lines per unit area.*

Potential and treatment	$\Delta\varrho$ (Ω cm)
Shear strains (HUNTER and NABARRO [14]) Born's approximation	$0.26 \cdot 10^{-20} N$
Same potential as HUNTER and NABARRO, exact solution [8]	$0.19 \cdot 10^{-20} N$
Including the line pole potential [8]	$0.685 \cdot 10^{-20} N$

by a factor of about three. This increase, however, is not nearly enough to account for the experimental data on copper and nickel which would require a further increase by a factor of about 40 (after allowing for the presence of edge dislocations, too) (*).

The main fact that has been neglected in the calculations mentioned so far is that dislocations in copper and nickel are extended into a pair of partial dislocations, containing a stacking fault ribbon between them. There is indeed rather strong experimental evidence that the main resistivity due to dislocations in metals like Cu and Ni does come from their stacking faults. Theoretical treatments [13, 15, 16] give an average reflexion coefficient of the stacking faults of about $\frac{1}{3}$, which does give the required increase by about a factor of 40 (+). Two quite different sets of experimental data were used for the comparison between experiment and theory and were found to give good agreement with experiment: a comparison between the electrical resistivity and the change in density or the stored energy due to dislocations [7, 8], and a determination of the resistivity of a dislocation line in copper based on single crystal measurements [17] and the theory of work-hardening [18].

4. – The Peierls stress in f.c.c. metals.

The Peierls stress in a (hypothetical) simple cubic lattice has been discussed by Professor FRANK. The quantity called σ_0 by him will be denoted τ_μ^0. It is the stress necessary to move forward by one lattice parameter a dislocation

(*) The comparison with experiment has been based [8] on the ratio of resistivity to change in density, which should be insensitive to the details of the model.

(+) If the stacking fault scattering is included the magnitude of the ratio of the resistivity to change in density comes out for dislocations in copper and nickel slightly higher than for vacancies.

line with dislocation strength b, which lies along one of the close-packed directions of the lattice. Fig. 5 illustrates this in a schematic way. It shows how such a dislocation line, when moving from one position of minimum energy to a neighbouring one, has to pass through a symmetric position of maximum energy, which is shown in Fig. 5.

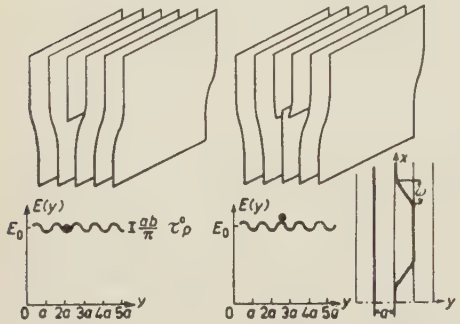


Fig. 5. – The Peierls stress τ_P^0 of a dislocation running in a close-packed direction (schematically). $E(y)$ = line energy of a dislocation line as a function of position. E_0 = mean line energy. w = width of a dislocation kink (of the order 100 Å).

At a finite temperature the dislocation line will contain a number of kinks, i.e. it will occasionally go over from one trough of potential energy to a neighbouring one. If a shear stress τ in the glide system of the dislocation is applied, there is a tendency to form more bulges going one way than the other. At a given temperature these bulges (or pairs of kinks) are formed with a certain frequency.

A maximum of the internal friction is obtained if the frequency of an applied alternating stress coincides with the frequency of bulge formation. This is a relaxation phenomenon similar to the Snoek effect of interstitial carbon atoms in b.c.c. iron discussed in ZENER's lectures.

It cannot be treated however with the conventional rate theory which arrives at Arrhenius' equation. A more complete treatment based on the theory of stochastic processes has been given by DONTH [19].

There is very strong evidence [20] that the relaxation phenomenon due to kink formation which we have been discussing is the so called BORDONI [21] relaxation peak which has been found in slightly cold-worked f.c.c. metals. Using Donth's form of the theory SEEGER, DONTH and PFAFF [22] have derived numerical values for τ_P^0 for a number of f.c.c. metals. The results for Cu and Al are given in Table II. They show that in these metals τ_P^0 is of the order of magnitude estimated by DIETZE [5]. It is however considerably larger (by a

TABLE II. – *The Peierls stress τ_P^0 of f.c.c. metals (G = shear modulus).*

Metal	τ_P^0/G	τ_P^0 [kg/mm²]
Cu	$4 \cdot 10^{-4}$	1.8
Al	$10 \cdot 10^{-4}$	2.7

factor of the order 10) than the critical shear stress of single crystals at liquid helium temperatures. This is not unreasonable since only a small fraction of the dislocation lines lies in a close packed direction and participates in the Bordoni relaxation peak. The rest of the dislocation lines averages out the Peierls stress and is able to move at very small stresses, even in the absence of thermal activation.

5. – Summary.

In this lecture we have treated a number of dislocation phenomena which cannot be adequately treated by the linear theory of elasticity.

In Sect. 2 it was shown that the influence of dislocations on the crystalline density can be accounted for in good agreement with experiment by allowing for the deviations from Hooke's law. The dilatation which accompanies a dislocation line is distributed over a rather large volume around the dislocation line.

The non-linear dilatation has a profound effect on the electrical properties of dislocations in metals. In Sect. 3 the theoretical and experimental situation concerning the electrical resistivity of dislocations is briefly reviewed. It is concluded that in Cu and Ni the dislocation resistivity is mainly due to the scattering of conduction electrons by stacking faults.

In Sect. 4 the results of a determination of the Peierls stress of metals from internal friction data (Bordoni relaxation) are reported.

Appendix (*)

In this Appendix we give a more detailed application of Zener's formula (equ. (10))

$$\Delta = \frac{1}{K}\left[\left(-\frac{\mathrm{d}\ln G}{\mathrm{d}\ln V} - 1\right) W_{\mathrm{sh}} + \left(-\frac{\mathrm{d}\ln K}{\mathrm{d}\ln V} - 1\right) W_{\mathrm{hyd}}\right], \tag{A1}$$

to the theory of dislocations, and a comparison with recent experiments. In eqn. (A1) Δ is the average dilatation in a self-strained isotropic elastic body, W_{sh} and W_{hyd} are the average energy densities due to shear stresses and hydrostatic pressure (as calculated from the linear theory of elasticity). G and K

(*) The author gratefully acknowledges valuable discussions on Zener's theory with Dr. P. Haasen and cand. phys. E. Mann. He wishes to thank Dr. W. Boas for making available to him before publication the experimental results of Clarebrough, Hargreaves and West [24].

denote the moduli of rigidity and compression. V is the specific volume of the crystal. In Zener's derivation of eqn. (A1) thermodynamics is used to relate the dilatation connected with shear to the change in the shear modulus due to a dilatation. The second term in eqn. (A1) is obtained by allowing for the change of compressibility with pressure and retaining linear terms in the pressure dependence of K only. Eqn. (A1) gives the same results as Murnaghan's [23] second order theory of finite deformation of an elastic body (E. MANN, to be published).

Eqn. (1) can be generalized for crystals. For cubic crystals we obtain

$$\text{(A2)} \qquad \Delta = \frac{1}{K}\left\{\left(-\frac{\mathrm{d}\ln c_{44}}{\mathrm{d}\ln V}-1\right)\cdot c_{44}\cdot 2\cdot(\varepsilon_{yz}^2+\varepsilon_{zx}^2+\varepsilon_{xy}^2)+\right.$$
$$+\left(-\frac{\mathrm{d}\ln (c_{11}-c_{12})/2}{\mathrm{d}\ln V}-1\right)\cdot\frac{c_{11}-c_{12}}{6}\cdot[(\varepsilon_{zz}-\varepsilon_{yy})^2+(\varepsilon_{xx}-\varepsilon_{zz})^2+(\varepsilon_{yy}-\varepsilon_{xx})^2]+$$
$$\left.+\left(-\frac{\mathrm{d}\ln (c_{11}+2c_{12})/3}{\mathrm{d}\ln V}-1\right)\cdot\frac{c_{11}+2c_{12}}{6}\cdot(\varepsilon_{xx}+\varepsilon_{yy}+\varepsilon_{zz})^2\right\}.$$

Here ε_{ik} are the components of the strain *tensor*; c_{ik} are Voigt's elastic constants, as usual.

The application of eqn. (A2) to straight dislocation lines in cubic crystals is straightforward, though rather tedious. We restrict ourselves to an isotropic medium with Poisson's constant ν. The total line energy E_{tot} of a screw dislocation in an isotropic medium is equal to the shear energy E_{sh}. We have

$$\text{(A3)} \qquad E_{\mathrm{sh}} = E_{\mathrm{tot}} = G\,\frac{b^2}{4\pi}\ln\frac{R}{r_0},$$

where b is the dislocation strength. R and r_0 denote the radii of cut-off. Part of the total line-energy of an edge dislocation

$$\text{(A4)} \qquad E_{\mathrm{tot}} = E_{\mathrm{sh}} + E_{\mathrm{hyd}} = \frac{G}{(1-\nu)}\cdot\frac{b^2}{4\pi}\cdot\ln\frac{R}{r_0}$$

is due to the energy of compression (and dilatation)

$$\text{(A5)} \qquad E_{\mathrm{hyd}} = \frac{(1-2\nu)(1+\nu)}{(1-\nu)\cdot 3}\,\frac{G}{1-\nu}\,\frac{b^2}{4\pi}\ln\frac{R}{r_0}.$$

The energy of shear is given by

$$\text{(A6)} \qquad E_{\mathrm{sh}} = \frac{2}{3}\,\frac{1-\nu+\nu^2}{1-\nu}\,\frac{G}{1-\nu}\,\frac{b^2}{4\pi}\ln\frac{R}{r_0}.$$

Inserting eqn. (A3) to (A6) into eqn. (A1) and using the relation

$$\text{(A7)} \qquad V\frac{\mathrm{d}p}{\mathrm{d}V} = -K$$

between hydrostatic pressure p, bulk modulus K, and volume, we obtain for

a screw dislocation

$$\frac{G\cdot\Delta}{W}=\frac{\mathrm{d}G}{\mathrm{d}p}-\frac{G}{K}, \tag{A8}$$

and for an edge dislocation

$$\frac{G\cdot\Delta}{W}=\left(\frac{\mathrm{d}G}{\mathrm{d}p}-\frac{G}{K}\right)\frac{2}{3}\frac{1-\nu+\nu^2}{1-\nu}+\frac{G}{K}\left(\frac{\mathrm{d}K}{\mathrm{d}p}-1\right)\cdot\frac{(1-2\nu)(1+\nu)}{3(1-\nu)}. \tag{A9}$$

Here W denotes the average density of total elastic energy.

In Table III we summarize a number of experimental determinations of the pressure dependence of the elastic constants of f.c.c. metals. We have included all the determinations of $\mathrm{d}G/\mathrm{d}p$ we could find in the literature; in view of the large number of measurements of $\mathrm{d}K/\mathrm{d}p$ we have confined ourselves to a selection of these.

As will be seen from Table III different determinations of the same quantity

TABLE III. – *Experimental values of the pressure dependence of elastic constants of f.c.c. metals, in units of* 10^{-6} cm²/kg$=10^{-6}$ bar^{-1}.

Metal	Author and technique	$\frac{\mathrm{d}\ln K/\mathrm{d}p}{\mathrm{d}\ln\frac{c_{11}+2c_{12}}{3}/\mathrm{d}p}$ or	$\frac{\mathrm{d}\ln G}{\mathrm{d}p}$	$\frac{\mathrm{d}\ln c_{44}}{\mathrm{d}p}$	$\frac{\mathrm{d}\ln\frac{c_{11}-c_{12}}{2}}{\mathrm{d}p}$
Cu polycrystalline	BIRCH (1937) torsional vibration	—	2.71	—	—
	BRIDGMAN (1935) static	3.65	—	—	—
Cu single crystal	LAZARUS (1949) supersonic pulses	2.84	—	1.13	2.45
Ni polycrystalline	BIRCH (1937) torsional vibration	—	1.82	—	—
	BRIDGMAN (1935) static	2.83	—	—	—
	BRIDGMAN (1929) static	—	1.8	—	—
Al polycrystalline	BIRCH (1937) torsional vibration	—	7.61	—	—
	BRIDGMAN (1935) static	5.83	—	—	—
	HUGHES and MAURETTE (1956) supersonic pulses	5.2	9.8	—	—
Al single crystal	LAZARUS (1949) supersonic pulses	5.47	—	7.49	—

References: F. BIRCH: *Journ. Appl. Phys.*, **8**, 129 (1937); P. W. BRIDGMAN: *Proc. Am. Acad.*, **64**, 39 (1929); P. W. BRIDGMAN: *Proc. Am. Acad.*, **70**, 285 (1935); D. S. HUGHES and CH. MAURETTE: *Journ. Appl. Phys.*, **27**, 1184 (1956); D. LAZARUS: *Phys. Rev.*, **76**, 545 (1949).

Misprints in Table V of LAZARUS (1949) have been corrected.

are usually in reasonably good agreement with each other. An exception is that $\mathrm{d}\ln G/\mathrm{d}p$ as determined by BIRCH on polycrystalline copper is considerably larger than the average of Lazarus' single crystal determinations. Since we were unable to decide which of the two measurements are more reliable we evaluated eqn. (A8) and (A9) for copper both by using Birch's and Bridgman's values and by using Lazarus' values, in the latter case employing as an average value $\mathrm{d}G/\mathrm{d}p = 0.8$.

For aluminium, where there is some scatter in the experimental data, we used Lazarus' single crystal data which are in quite good agreement with Birch's and Bridgman's polycrystalline data.

In Table IV we give (in the form of the dimensionless quantity $G\cdot\Delta/W$) the ratio of the average dilatation and the volume density of energy due to screw and edge dislocations as calculated for Cu, Ni and Al from eqn. (A8) and eqn. (A9). From these quantities the ratio between the fractional change in density $\Delta D/D$ and the dislocation energy per unit mass S has been computed. $\Delta D/DS$ has been determined experimentally from the fractional increase in density and the stored energy released during the recrystallization of polycrystalline copper and nickel. Table IV gives the experimental results

TABLE IV. – *Comparison between experiment and theory.*

	$(G\cdot\Delta)/W$		$\Delta D/DS$ [10^{-4} g/cal]		
	screw dislocations	edge dislocations	screw dislocations	edge dislocations	experimental
Cu					
LAZARUS	0.5	0.6	4.2	5.0	9.6 (30% compression) 10.8 (55% compression)
BIRCH, BRIDGMAN	0.95	1.03	7.9	8.6	13.1 (70% compression)
Ni	1.0	1.2	4.9	5.8	7.5 ($nd/l = 1.41$) 7.6 ($nd/l = 2.34$)
Al	1.6	1.45	7.5	7.0	no experimental values available

of CLAREBROUGH, HARGREAVES and WEST on copper compressed at room temperature [24] and on nickel deformed in torsion at room temperature [25] (n = number of full turns of the specimen, d — diameter of the specimen, l = length of the specimen).

Before comparing the experimental data with the theory we have to discuss certain corrections that have to be applied to eqn. (A8) and (A9). They are due to the fact that Zener's relation eqn. (A1) ceases to be valid in the center of the dislocation lines. We may say that we should use different inner radii of cut-off for the dilatation and the dislocation energy. In the case of Cu and Ni a further correction has to be applied since the preceding treatment does not allow for the extension of the dislocations into partials. We estimate that these two corrections give an increase of $\Delta D/DS$ of about 15% to 20%.

Taking into account these corrections we obtain rather good agreement between experiment and theory for nickel. For copper there is order of magnitude agreement, although the theoretical values appear to be somewhat too low. We think that this slight discrepancy may be due to vacancies created during cold-work contributing to the change in density and stored energy during recrystallization [8]. Assuming that a vacancy in copper has an energy of formation of 0.9 eV and expands the volume of the crystal by one atomic volume it is found that for vacancies $\Delta D/DS = 30.4 \cdot 10^{-4}$ g/cal. A small contribution of vacancies to the stored energy liberated during recrystallization (of the order of $10\% \div 20\%$) would suffice to bring theory and experiment into quantitative agreement.

REFERENCES

[1] H. B. Huntington, J. E. Dickey and R. Thomson: *Phys. Rev.*, **100**, 1117 (1955).
[2] G. Colonnetti: *Atti Acc. Naz. Lincei, Rend., Cl. sci. fis. mat. natur.*, (5) **272**, 155 (1918).
[3] C. Zener: *Trans. Amer. Inst. Min. Met. Eng.*, **147**, 361 (1942).
[4] R. Landauer: *Phys. Rev.*, **82**, 520 (1951).
[5] H. D. Dietze: *Zeits. f. Phys.*, **131**, 156 (1952); **132**, 107 (1952).
[6] H. Stehle and A. Seeger: *Zeits. f. Phys.*, **146**, 217 (1956).
[7] L. M. Clarebrough, M. E. Hargreaves and G. W. West: *Proc. Roy. Soc. London*, A **232**, 252 (1955).
[8] A. Seeger and H. Stehle: *Zeits. f. Phys.*, **146**, 262 (1956).
[9] C. Crussard and F. Aubertin: *Rev. Metallurgie*, **46**, 354 (1949).
[10] A. Seeger: *Z. Naturforsch.*, **11***a*, 724 (1956); *Acta Met.*, **5**, 24 (1957).
[11] W. W. Beeman, P. Kaesberg, J. W. Anderegg, and M. B. Webb: *Handbuch der Physik*, **32**, p. 441.
[12] P. Hirsch: private communication.
[13] A. Seeger: *Can. Journ. Phys.*, **34**, 1219 (1956).
[14] S. C. Hunter and F. R. N. Nabarro: *Proc. Roy. Soc. London*, A **220**, 542 (1953).
[15] P. G. Klemens: *Can. Journ. Phys.*, **34**, 1212 (1956).
[16] H. Stehle: *Thesis*, Stuttgart, 1957.
[17] T. H. Blewitt, R. R. Coltman and J. K. Redman: *Rep. Conf. Defects in Solids*, The Physical Society, London (1955), p. 369.
[18] A. Seeger: *Kristallplastizität*, in *Handb. d. Phys.*, Vol. **7/2**.
[19] H. Donth: *Thesis Technische Hochschule*, Stuttgart, 1957; *Z. Phys.* **149**, 111 (1957).
[20] A. Seeger: *Phil. Mag.*, **1**, 651 (1956).
[21] P. G. Bordoni: *Ric. Sci.*, **19**, 851 (1949); *Journ. Ac. Soc. Am.*, **26**, 495 (1954).
[22] A. Seeger, H. Donth and F. Pfaff: *Disc. Faraday Soc.*, **23**, 19 (1957).
[23] F. D. Murnaghan: *Finite Deformation of an Elastic Solid* (New York, 1951).
[24] L. M. Clarebrough, M. E. Hargreaves and G. W. West: *Acta Met.*, **5**, 738 (1957).
[25] L. M. Clarebrough, M. E. Hargreaves and G. W. West: *Phil. Mag.*, **1**, 528 (1956).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Dislocations in Germanium and Silicon.

H. G. VAN BUEREN, J. HORNSTRA and P. PENNING

Philips Research Laboratories - Eindhoven

Introduction.

This paper describes work done recently in Eindhoven on the subject of dislocations in the diamond structure and on the plastic deformation of its most important representative—germanium. The work consists of three parts: Sect. **1** discusses the nature of dislocations in germanium and silicon; Sect. **2** is devoted to the formation of dislocations by thermal stresses and Sect. **3** discusses several more general aspects of the plastic deformation of germanium.

1. – Geometry of dislocations.

From a variety of observations it is known that plastic glide in germanium and silicon occurs on {111} planes only, whereas the glide direction is invariably ⟨110⟩ [1]. This is not surprising since the spacing between neighbouring lattice planes in the diamond lattice (Fig. 1) is maximum for {111} planes, and since the shortest lattice vector is indeed a vector along ⟨110⟩, of magnitude $\frac{1}{2}a\sqrt{2}$, where a is the edge of the cubic unit cell. Evidently all dislocations responsible for plastic flow have such a Burgers vector and a {111} plane as glide plane. As any direction in the crystal may be regarded as being composed of steps in various ⟨110⟩ directions, a systematic study of dislocations in the diamond lattice calls primarily for a consideration of the various possible combinations of ⟨110⟩ line and ⟨110⟩ Burgers vector directions (not necessarily within the same {111} plane) [2].

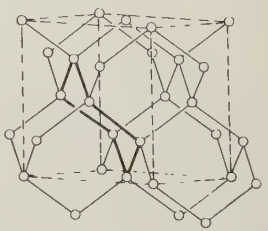
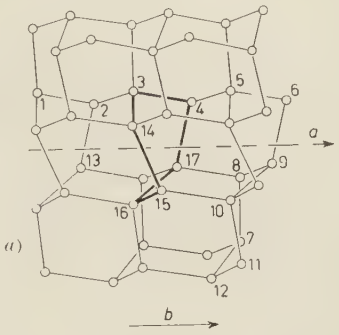
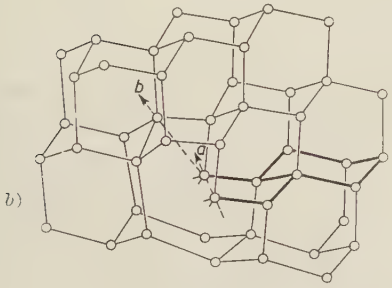


Fig. 1. – The diamond structure.

There are three such combinations possible, leading to the so-called *simple* dislocations, viz. a screw dislocation (Fig. 2*a*), an edge dislocation (Fig. 2*b*) and a so-called 60° dislocation (Fig. 2*c*); the latter is predominantly, but not fully, edge type.

The latter two types of dislocation are characterized by the occurrence of free or « dangling » bonds along the axis. These are absent in the screw dislocation. However, what remains in all three types is an appreciable rotation of the bond directions of the atoms near the dislocation axis.





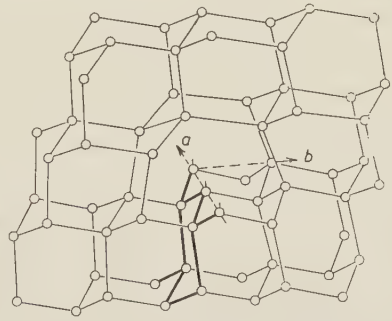


Fig. 2. – Simple dislocations in the diamond structure (dislocations with axis and Burgers vector both along ⟨110⟩). *a*) screw dislocation; *b*) edge dislocation (glide plane {001}); *c*) 60°-dislocation (glide plane {111}).

Both factors are very important in determining the structure and properties of a dislocation in a homopolar crystal. To the effect of free bonds on the motion of a dislocation we shall return in Sect. **3**.

In general there will exist a tendency for rearrangement of the atomic structure near the axis of the dislocation (core structure) when this leads to a decrease, either of the number of dangling bonds or of the deviations of the bond directions from the normal tetrahedral orientation. Mostly, these two influences counteract each other and in the absence of quantitative knowledge about the corresponding changes in free energy, it cannot yet be predicted which of the possible arrangements will be stable.

A possible rearrangement of the screw dislocation core is depicted in Fig. 3*a*. It is derived from the structure of Fig. 2*a* by breaking the bonds 2-3, 4-5, etc., thereby leaving dangling bonds at 1, 2, 3, 4, 5, etc. This is equivalent to a kind of splitting of the dislocation into two components, resulting in dimi-

nished deviations of the bond directions, but also in the formation of dangling bonds. Free bonds on two neighbouring atoms such as 1 and 2, 3 and 4, etc., can, however, possibly interact and form a *double* bond between these atoms.

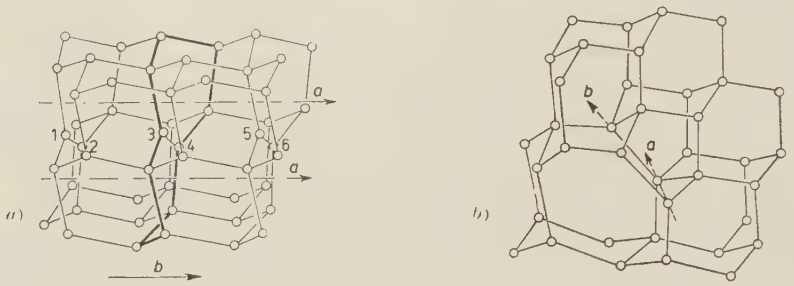


Fig. 3. – *a*) Screw dislocation of Fig. 2 *a*), with rearranged core structure. The dangling valencies of atoms 1 and 2, 3 and 4, 5 and 6, etc., might form double bonds. *b*) Edge dislocation of Fig. 2 *b*) with rearranged core structure. All free bonds have disappeared.

The edge dislocation structure of Fig. 2*b* can be rearranged by removing all the atoms with the broken bonds. Then a structure occurs (Fig. 3*b*) in which no broken bonds at all exist, but instead of normal 6-rings, 5- and 7-rings occur. Whether this or the reverse transition in the structure of an edge dislocation will actually occur, depends among other things on the possibility of removing (e.g. by diffusion) the superfluous atoms, and vice versa. Possibly this transition takes place at high temperatures only.

A type of dislocation with Burgers vector along ⟨110⟩ and a different axis, not ⟨110⟩, can be obtained by the superposition of two simple dislocations. In this way six new dislocations can be constructed, two along ⟨100⟩ and four along ⟨211⟩, one of the latter being an edge dislocation with {111} glide plane. A description of these has been published elsewhere [2]. Most of these dislocations can exist, at least theoretically, in various forms, differing in the number of broken bonds per unit length. Only one of them, an edge dislocation along ⟨100⟩ (with a Burgers vector of course again along ⟨110⟩) may be mentioned here, as it has been observed as a constituent of small-angle grain boundaries in germanium. Its structure is depicted in Fig. 4. By the removal of the atoms indicated by an arrow, a structure without broken bonds can be formed.

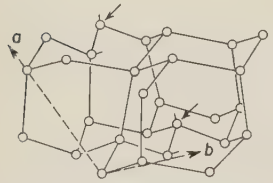


Fig. 4. – Edge dislocation along ⟨100⟩ (with Burgers vector along ⟨011⟩, glide plane $\{0\bar{1}1\}$).

It will be clear from the foregoing that in the diamond structure edge dislocations *without* any broken bonds are theoretically very well pos-

sible, although generally these will be less mobile under the action of a shear stress.

The splitting up of dislocations into partial dislocations is possible in a great many ways, in more ways even than in the related face-centred cubic structure, owing to the existence of two types of atom sites in the diamond lattice. The splitting can be studied with the aid of the analysis made for the face-centred cubic structure by FRANK and NICHOLAS [3]. A simple dislocation along a $\langle 110 \rangle$ direction can split up into partials with the connecting stacking fault either along the glide plane or along the other $\{111\}$ plane through its axis. The latter possibility can, however, easily be transferred into the former [2]. We shall only illustrate one particular case here, viz. the splitting of a 60° dislocation into two partials with a stacking fault in the glide plane. Three types of partial dislocations bounding this stacking fault, with Burgers vectors along $\langle 211 \rangle$ directions, are possible (Fig. 5*a*). They are formed by the rotation of a « ribbon » of atoms in the core of the original dislocation (Fig. 5*b*). Motion of a partial dislocation includes successive rotation of neighbouring atom ribbons. Extended dislocations are therefore possibly less mobile than perfect dislocations.

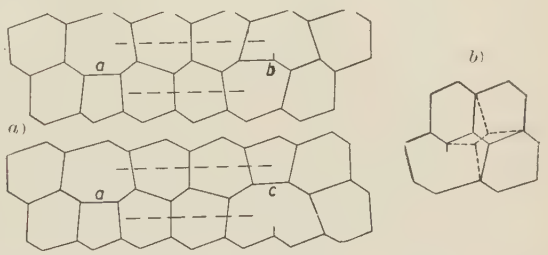


Fig. 5. – *a*) Two possible ways in which a 60° dislocation can be split into two partial dislocations and a stacking fault in the $\{111\}$ glide plane. The lattice is viewed in projection on a $\{110\}$ plane. There are three types of bounding partials, denoted by *a*, *b* and *c* respectively. *b*) The formation of a partial dislocation by the rotation of a ribbon of atoms in the core of the original dislocation.

It is well-known that growth twins are easily formed in germanium and silicon. If from this the conclusion could be drawn that the energy of a coherent twin boundary is relatively small, the formation of extended dislocations with stacking faults should be energetically favourable, as no increase of the number of broken bonds, and generally a decrease of the rotations of the bond directions, are connected with it.

2. – Thermal strain and dislocations in germanium.

From the study of the electrical properties of germanium and silicon crystals prepared by the Czochralski-method, it is known that these properties are largely determined by the rate of growing and of cooling of the crystal. Since the use of the etch pit technique to indicate dislocations, it has been shown

that, apart from the impurity content, also the dislocation density in the crystals depends sensitively on the way of preparation. Dislocations, therefore, might be connected with the variations in the electrical properties. These dislocations are formed during the cooling of the crystal drawn from the melt, by the action of thermal stresses arising from the inhomogeneous temperature distribution in the solidified crystal. These produce some plastic deformation. In this section, we shall outline the results of some experiments on the nature of this deformation.

The thermal stresses arise to oppose the deformation by the inhomogeneous thermal expansion. When a germanium crystal is quenched from a temperature of 900 °C into a bath of liquid tin (400 °C), the stresses reach values of 50 kg/mm². Under these stresses and at this relatively low temperature the material cracks rather than flows plastically. This has in fact been observed [4].

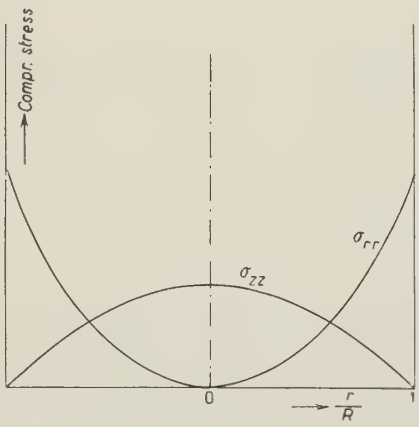


Fig. 6. – Parabolic stress distribution in a radially cooling cylinder of radius R. σ_{zz} axial compressive stress. σ_{rr} radial compressive stress (hydrostatic stresses have been subtracted).

On the other hand, on radial cooling in a gas atmosphere the stresses generally remain below a few kg/mm²; moreover, the temperature decreases but slowly and the crystal can now easily yield to this stress by plastic deformation (see also Sect. **3**). The thermal strains reach their maximum a short time after the cooling has started and afterwards decay, together with the temperature, exponentially. The distribution of stress through a perfectly elastic cylindrical specimen in which the heat flow is radial, remains then parabolic and is depicted in Fig. 6, after substracting the hydrostatic component (that can never give slip). One is left with the compressional stresses:

$$\sigma_{zz} = -\left(1 - \frac{r^2}{R^2}\right)\sigma_0 \quad \text{in the axial direction,} \tag{2.1}$$

and

$$\sigma_{rr} = -2\,\frac{r^2}{R^2}\,\sigma_0 \quad \text{in the radial direction,} \tag{2.2}$$

were σ_0 depends on cooling rate, expansion coefficient etc. and decays with time. The net stress is thus a radial compression at the surface and an

axial compression near the axis of the crystal rod. Efforts have been made [4, 5] to determine the geometry of the plastic flow caused by this initial elastic stress distribution. For this purpose, it has been assumed in [4] that only dislocations on {111} planes with ⟨110⟩ Burgers vectors become active. Further, the number of dislocations then formed is taken proportional to the magnitude of the shear stresses in the slip plane. As there are three slip systems per slip plane, the number of dislocations in a given slip plane is then proportional to the sum of the three slips on these systems. It is this quantity that has been computed.

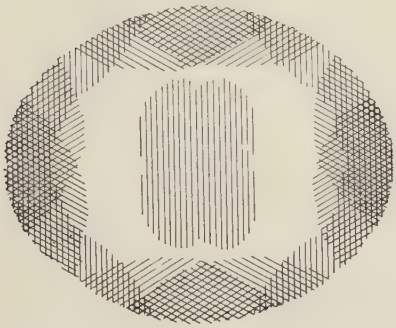


Fig. 7. – Predicted line-pattern of etch pits on a {111} cross-section of a cylindrical ⟨110⟩ germanium crystal, cooled in a gas atmosphere in such a way that radial heat flow occurs only. Basic assumption: plastic strain reflects initial elastic stress distribution.

The results of such a computation are presented in Fig. 7. The shaded areas are parts of the elliptical {111} cross-section of a cylindrical crystal with its axis along a ⟨110⟩ direction, on which the slip on the slip planes indicated by the direction of the shading, exceeds a certain value.

Experimentally, dislocations can be indicated by etching techniques. It is generally believed that by suitable etching (e.g. CP-4 etch) nearly all dislocations intersecting with the surface give rise to the formation of an etch pit. It is therefore to be expected that in thermally quenched germanium, too, etch pits will appear along rows parallel to the tracing of the active slip planes. That is, by etching a {111} face of a ⟨110⟩ crystal, rows of etch pits should occur that are distributed according to the theoretical pattern of Fig. 7. This is, indeed, the case, as is illustrated in Fig. 8.

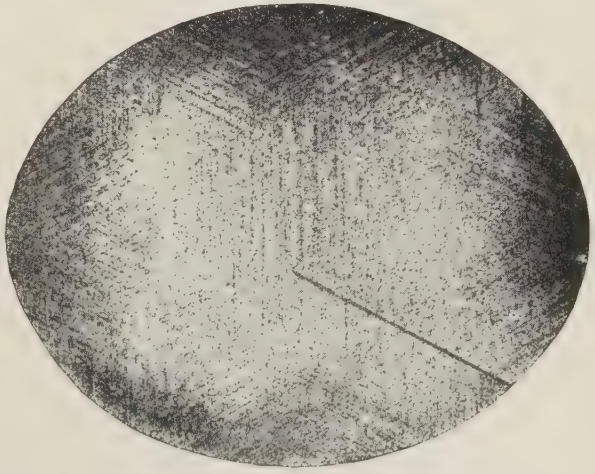


Fig. 8. – Observed pattern of etch pits on a {111} cross-section of a cylindrical ⟨110⟩ crystal cooled in radial heat flow.

The fair agreement

between observed and predicted etch pit pattern rests primarily on the assumption, made in the computation, that the amount of plastic flow that has taken place reflects fully the initial elastic thermal stress distribution. This is only the case when but a small constant fraction of the thermal stresses has been released by plastic flow, so that the remaining elastic strain distribution conforms approximately to the theoretical distribution in a non-plastic material. It is not *a priori* certain that this condition should hold. A situation that can be regarded as the extreme opposite is one where the stress release by plastic flow has been complete. The theoretical pattern of etch pits has also been derived for this hypothetical case; it is depicted schematically in Fig. 9. The agreement with the observed distribution in Fig. 8 is not nearly as good as in the former case, especially as regards the form of the central etch pit density maximum and minimum zones. The second hypothesis is thus evidently to be rejected and the actual situation conforms much better to the hypothesis of little plastic flow.

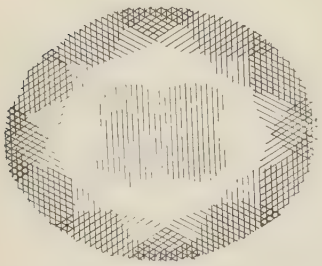


Fig. 9. – Predicted pattern of etch pits in the system of Fig. 7 and 8, on the assumption of complete stress release by plastic flow.

One can ask why the thermal stresses are only partly released, as the maximum stresses occur at temperatures above 800 °C, where germanium is a very plastic material.

The answer probably lies in the long distance the dislocations have to move in order to take up all thermal strain, or in the large number of dislocations that have to be created for this purpose. The strain amounts in actual cases—maximum temperature difference of about 60 °C between centre and boundary [4]—to about $4 \cdot 10^{-4}$. To accommodate this, a total dislocation length N cm per cm^3 would have to move a distance s, where $Ns \approx 10^4$ cm^{-1}. Actual dislocation densities observed after gas quenching are of the order of 10^5 cm^{-2}, from which it should follow that $s \approx 0.1$ cm. Dislocation migration over such long distances in the little time available (about 1 second) seems improbable. From the observed temperature dependence of plasticity, to be outlined in Sect. 3, it can be inferred that the maximum strain that can be produced by this number of dislocations is 10^{-5}. More than 10^6 dislocations per cm^2 would be needed to accommodate *all* thermal stress. Such high dislocation densities have never been observed after quenching; the dislocation multiplication process (a Frank-Read mechanism), is too slow for this. Presumably, therefore, the thermal stresses can not be released by more than a small fraction.

3. – Plastic flow of germanium and silicon.

3·1. *Experiments.* – In the previous section the subject of plastic flow in germanium has repeatedly been touched upon. We shall now discuss some more general features of this problem; we shall limit ourselves to the mechanical aspects only and not discuss the electrical and microscopical changes that occur on plastic deformation, however important they may be.

Homopolar substances are generally brittle and germanium and silicon form no exception; they are glass-brittle at room temperature and fracture by cleavage when the tension exceeds 10 to 20 kg/mm². However, in 1952 it was discovered that above 400 °C germanium single crystal rods would creep plastically under an applied load corresponding to a tensile stress of a few kg/mm². Creep strains up to 10% could be obtained without fracture, given enough time for the crystal to yield. Crystallographic slip on the $\{111\}\ 110$ systems was observed microscopically. Under specific loading conditions also mechanical twinning was found to contribute to the deformation process. We shall not discuss this complication in this paper.

At higher temperatures the ductility of germanium increases further and instantaneous flow, no longer characterized by special yield phenomena, occurs above 600 °C.

Silicon shows analogous behaviour; in this element plastic creep becomes observable above 600 °C [6] (the melting point is 1420 °C instead of 937 °C for Ge). As reliable data on silicon are still scarce, we shall restrict ourselves in the following to a discussion of the properties of germanium only.

Quantitative investigations of the plasticity of germanium (and silicon as well) can, for purposes of presentation, be divided into three categories.

The first category concerns the afore-mentioned creep experiments at relatively low temperatures (below 550 °C in Ge). The loading of germanium single crystals with a constant force results in plastic flow that increases with time. This creep phenomenon is characterized by an *initial period*, during which no strain at all or only very little strain occurs; a *steady state* period, during which the creep rate is fairly constant and most of the strain is produced; and a *work-hardening* period, within which the strain rate decreases again (Fig. 10). The actual shape of the creep curve depends on the manner of loading, the crystallographic orientation and the impurity content. In bending tests a sharply defined *delay time* has been repeatedly observed, during which no strain is detected at all. Such a delay time is usually also observed in uniaxial loading; however, it is then under comparable conditions of stress and temperature much shorter and often even entirely absent. In the latter case, the strain during the initial period is found to increase from the very beginning of the test, at a slow but nevertheless observable constant

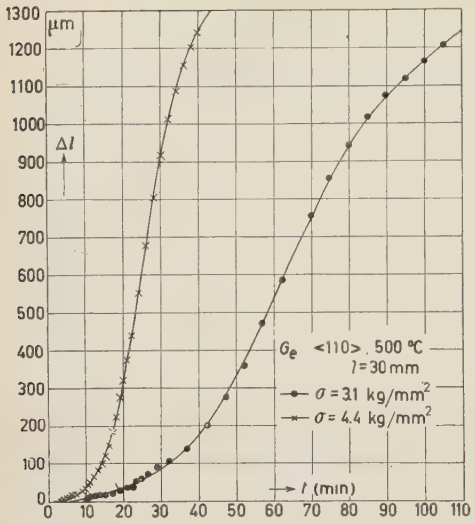


Fig. 10. – Typical creep curves of germanium deformed at a load of 3.1 and 4.4 kg/mm² respectively at 500 °C. The delay periods amount to 2 and 6 minutes in these two cases. In the region of extension between about 250 and 1000 μm, the creep rate is approximately constant (steady state period). After this, work-hardening begins.

rate, until after a fairly well determined time interval the steady state, characterized by a different creep rate, is obtained. The reason for these differences is probably to be sought in the inhomogeneity of stress in a bent specimen (a large stress exists only in the outer layers). During a tension or compression test the stress distribution is much more homogeneous, although not completely so, as is often assumed; the homogeneity depends sensitively on the state of the crystal surfaces (cf. the recent photoelastic study on extended AgCl bars by NYE [7]).

The length of the delay time and of the initial period respectively, the creep rate during the latter and the steady state creep rate depend sensitively on the temperature and the stress. The temperature dependence of all these quantities can, in the region of medium stress (about 4 kg/mm²), be very well described by an activation energy of about 1.7 eV. To the form of the stress dependence we shall return in the next section.

The second category of plasticity experiments on germanium concerns deforming at a constant strain rate in the temperature region between about 500 °C and 700 °C. A sharp *upper yield-point* for plastic flow can be observed under these conditions. The yield stress depends inter alia on the rate of loading and on the temperature; it is not very sharply defined. A typical stress-strain curve is depicted in Fig. 11. After yielding, plastic glide occurs, accompanied by work-hardening, the coefficient of which again depends strongly on temperature. The transition between the « creep category » and the « flow category » rests actually on the latter property: below 550 °C the work-hardening coefficient becomes practically infinite.

Above 700 °C the specific features discussed above are no longer observable in practical tests and germanium has become a plastic material (third category); at 900 °C a germanium bar deforms easily under its own weight. Quantitative data on this behaviour are not available.

One further item concerning the plasticity of germanium should be men-

tioned here. The dislocation density during the deformation process can easily be studied by the etch pit technique. It is then found that, both in creep and in high temperature experiments, the density of etch pits, and hence of dislocations, increases very rapidly from its original value (10^3 to 10^4 cm^{-2} in most cases) to a few times 10^7 cm^{-2} during the first two or three percents of strain, to remain approximately constant afterwards. The onset of the saturation coincides more or less with the beginning of the work-hardening region in creep tests. During the delay period or during the initial period of the creep curve, no or very few slip lines can be observed on the surface of the crystals; these appear in large numbers only during steady state creep.

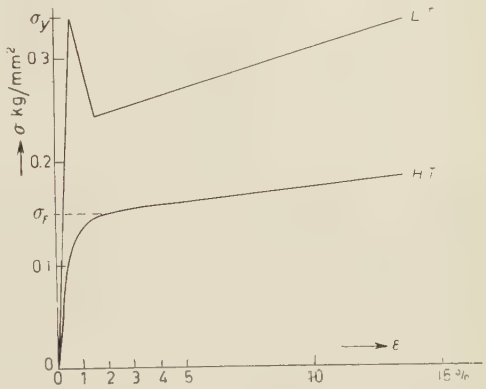


Fig. 11. – Schematic stress strain curves of a germanium single crystal deformed at a constant strain rate at a relatively low temperature (550 °C, LT) or at a relatively high temperature (650 °C, HT). σ_Y: yield stress; σ_F: flow stress.

Actual etch pit countings during the initial period show that the overall dislocation density does not increase in this period. Only local regions of a higher dislocation density appear, indicating an inhomogeneity of the initial deformation.

Along with our own work, use has been made of the published work of Gallagher [8], Vogel [9], Greiner [1], Treuting [1], Carreker [10] and Patel and Alexander [11] to arrive at the observational conclusions discussed in this section.

3·2. *Theory.*

3·2.1. Impurity locking theory. – The occurrence of a delay time together with an upper yield point as characteristics of the plastic deformation of germanium have led to the widely accepted view that the initial stages of flow are governed by the interaction of dislocations with impurity atoms. As no correlation has been found with the electrical properties (carrier density), those impurity atoms interacting with the dislocations are either present in such very minute amounts to remain undetected, or they are electrically inactive, e.g. oxygen or carbon. To explain the observed stress and temperature dependence, a theory proposed by Fisher [12] and based on the locking of dislocations by impurities would be expected to apply. According to this

theory, under the action of a shear stress σ dislocations are pulled loose from the impurity atoms blocking them. The energy to be supplied for this process reaches a maximum value as a function of the length l of dislocation that has broken loose; no more energy is needed to extend the free length further. The maximum energy is

$$W_0 \approx 2\sqrt{2}(1-\Delta)^{\frac{3}{2}} \frac{\gamma_0^2}{b\sigma}, \tag{3.1}$$

where γ_0 stands for the line energy (about Gb^2, G is the shear modulus) of the free dislocation, b is the Burgers vector and $\Delta \cdot \gamma_0$ is the binding energy of impurity atoms, measured per unit length of the blocked dislocation line.

FISHER considered W_0 as the activation energy for the action of dislocation sources. The delay or initial period should therefore depend on stress and temperature through the term:

$$\exp\left[\frac{W_0}{kT}\right] = \exp\left[2\sqrt{2}(1-\Delta)^{\frac{3}{2}} \frac{\gamma_0^2}{bk} \frac{1}{\sigma T}\right], \tag{3.2}$$

thus through the product (σT) only. Inversely, the upper yield stress in tests of the second category should be related to the temperature by

$$\sigma_y \cdot T = \text{const}. \tag{3.3}$$

Recently, COTTRELL [13] has criticized Fisher's theory on the ground that the concept of line tension cannot be applied to very short dislocation segments. The change in energy of the *core* of a dislocation when it pulls loose from its impurity atoms should rather be considered instead. The activation energy for the action of a Frank-Read source is then found as:

$$W_0 = A\left(1 - \frac{\sigma}{\sigma_0}\right)^3, \tag{3.4}$$

and (3.3) must be replaced by:

$$\sigma_y = \sigma_0 - \text{const}\ T^{\frac{1}{2}}. \tag{3.5}$$

Actually, the stress dependence of the delay period etc., and the temperature dependence of the yield stress have not yet been very accurately determined, owing to a large spread in the experimental results (probably caused by stress inhomogeneity due to surface effects). However, from our own experiments it follows clearly that the characteristic parameters of the initial creep period depend on stress much less sensitively than is suggested by for-

mula (3.2), in which stress and temperature influences are comparable. From the *temperature* dependence an activation energy can be determined with fair accuracy. This is found to be 39 000 to 41 000 cal/mole (1.7÷1.8 eV) at $\sigma = 4$ kg/mm^2. When tried to express the observed *stress* dependence by a relation of the form $\exp[A/\sigma]$, the activation energy as derived from A ($5 \cdot 10^8$ dyn/cm^2) was found to be only 2 000 cal/mole, thus very much less.

Moreover, the value of A to be used in (3.2) to explain this relatively weak stress dependence would correspond to a binding energy of an impurity atom to a dislocation of about 0.003 eV (per atomic spacing) only, which also seems far too small. For these reasons we believe that Fisher's theory does not apply here. Either Cottrell's argument must be used or impurity locking is not the rate-governing process of dislocation generation in germanium. Our own observations on the length of the delay time as a function of the stress can be described satisfactorily with the aid of formula (3.4); a correlation with the impurity content of the material has also been established without doubt. It is not possible, however, to decide unambiguously between the two possibilities.

The impurity locking theory is, of course, not able to describe more than the initial stages of creep and yield, viz. the *generation* of the dislocations. On the subsequent behaviour of the dislocations, thus on the steady state creep rate ultimately reached, it has nothing to say. This is also observed to be strongly stress dependent, however, equally strongly as the length of the initial period. An explanation of this in terms of impurity-locking is difficult.

3·2.2. Impeded motion theory. – The mechanism underlying the specific plastic properties of germanium and silicon could, in principle, be very well related to the special properties of dislocations in the diamond lattice as such, namely the occurrence of free valencies along the dislocation line, or the relative immobility of extended dislocations and of dislocations with « rearranged core structures » (see Sect. 1). To move a dislocation of arbitrary type in the diamond lattice, bonds must be continuously broken and remade, or bond directions or whole atom ribbons must be successively rotated. These processes more or less impede the *motion* of dislocations under action of a stress. The situation can be formally described by assigning to dislocations in the diamond structure a high *localized* Peierls force, localized in the sense that only in discrete parts of the dislocation line the mobility is severely reduced.

Moreover, in the absence of a force on a dislocation, this imperfection may spontaneously transform into a relatively immobile structure, e.g. by the combination of every two free valencies to a double bond (comp. Fig. 3*a*), or by a sort of climb process (Fig. 3*b*). The latter is temperature activated,

the former process hardly or not at all. Hence, also the *generation* of dislocations may be obstructed by the specific binding forces between the germanium atoms only, without impurity atoms playing a part.

A formal theory of plastic flow in germanium could now be the following. We assume that along the dislocation axis sites occur, with mutual spacing d, that form obstacles to motion (e.g. broken bonds) and that an energy E must be supplied to move these sites over one atomic distance (to break the bond) in the absence of a force. The corresponding energy of activation when a shear stress σ, producing a force σb per unit length on the dislocation, is present, is then simply given by

$$E - \sigma b^2 d \,. \tag{3.6}$$

For reverse motion the activation energy is

$$E + \sigma b^2 d \,. \tag{3.6b}$$

That is, under these assumptions the stress and temperature dependence of the creep characteristics (creep rate, inverse initial time, etc.) of germanium would be governed by a term of the form

$$\nu_0 \left\{ \exp\left[-\frac{E - \sigma b^2 d}{kT} \right] - \exp\left[-\frac{E + \sigma b^2 d}{kT} \right] \right\} = 2\nu_0 \exp\left[-\frac{E}{kT} \right] \sinh\left(\frac{\sigma b^2 d}{kT} \right) , \tag{3.7}$$

where ν_0 ($\approx 10^{13}\ \mathrm{s}^{-1}$) is the atomic vibration frequency. It is of course not *a priori* certain that the initial creep stages (dislocation generation) are described by the same values of the parameters E and d as the later stages (steady creep).

Interpretation of our own creep observations along these lines proved very well possible. The length of the initial period and the rate of creep during this period both showed the same temperature and stress dependence, which could be formally described by putting in (3.7)

$$E = 44\,000 \text{ cal/mole } (= 1.9 \text{ eV}) ,$$

$$d = 4 \cdot 10^{-7} \text{ cm} .$$

In view of the uncertainty of the determination of the stress dependence, this value of d seems reasonable when it is remembered that free valencies occur generally at a few atomic distances or more from each other along the

dislocation axis, depending on the crystallographic orientation of the latter. As to the value of E, not much can be said at present.

The rate of steady creep could also be described by a corresponding formula, with about the same parameter values. In general, the creep curve seems to depend on the stress *as a whole* in the same manner.

In the impeded motion theory it is easily found that the steady state creep rate of germanium must be expressed as

$$\dot{\varepsilon}_s = 2Nb^2\nu_0 \exp\left[-\frac{E}{kT}\right] \sinh\left(\frac{\sigma b^2 d}{kT}\right). \tag{3.8}$$

Here N is the dislocation density, assumed to be constant. Inserting the values quoted above, $\dot{\varepsilon}_s$ is found to be $2\cdot10^{-6}$ s^{-1} for $N=10^7$ cm^{-2}, $T=470$ °C and $\sigma = 4$ kg/mm^2. Observed creep rates and etch pit densities under these circumstances scatter around 10^{-6} s^{-1} and 10^7 cm^{-2} respectively.

The multiplication of dislocations has been optically observed [14] to occur by the Frank-Read process. In order to emit a dislocation, the source dislocation has to run a distance of the order of the length l of the source. Thus each « obstacle » on it must carry out l/b jumps, and a relaxation time for dislocation multiplication may be defined:

$$\tau = \frac{l}{b}\nu_0^{-1} \exp\left(\frac{E}{kT}\right) \frac{1}{\sinh\left(\sigma b^2 d/kT\right)}. \tag{3.9}$$

Taking $l/b = 10^4$ and again $\sigma = 4$ kg/mm^2, $T = 470$ °C, τ follows as a few minutes. The multiplication process is thus *slow*, and appreciable initial periods have to be expected in creep tests.

The fact that this slowness of the multiplication process also plays an important role during thermal quenching (see Sect. 2), becomes apparent when it is realized from (3.9) that at 800 °C $\tau = 0.1$ s, at 900 °C $\tau = 0.01$ s. During the first, most important, second of the quench the dislocation density can thus not rise by more than a factor 10 to 100.

Recently, Haasen [15] has tried to explain the temperature dependence of the flow stress after yielding (in experiments of the second and third category) in a somewhat different way, namely in terms of the diffusive motion that will occur in the dislocation core when the imperfection moves over its glide plane (there is a kind of crack in the core of a 60° dislocation, Fig. 2*c*). The activation energy describing this temperature influence, about 1.300 cal/mole, should then be the energy that must be supplied to the dislocation in order that its broken bond structure can *diffuse* along together with it.

REFERENCES

[1] E. S. Greiner: *Journ. Metals*, **7**, 283 (1955); R. G. Treuting: *Journ. Metals*, **7**, 1027 (1955).
[2] J. Hornstra: *J. Phys. Chem. Solids* (to be published).
[3] F. C. Frank and J. F. Nicholas: *Phil. Mag.*, **44**, 1213 (1953).
[4] P. Penning: *Philips Research Reports* (to be published April 1958).
[5] E. Billig: *Proc. Roy. Soc.*, A **235**, 37 (1956).
[6] G. L. Pearson, W. T. Read and W. L. Feldmann: *Acta Metall.*, **5**, 181 (1957).
[7] J. F. Nye, R. D. Spence and M. T. Sprackling: *Phil. Mag.*, **2**, 772 (1957).
[8] C. J. Gallagher: *Phys. Rev.*, **88**, 721 (1952); **92**, 846 (1953).
[9] F. L. Vogel: *Acta Metall.*, **3**, 95 (1955); *Journ. Metals* (*Transact. Suppl*), **8**, 946 (1956).
[10] R. P. Carreker: *Journ. Metals* (*Transact Suppl.*), **8**, 111 (1956).
[11] J. R. Patel and B. H. Alexander: *Acta Metall.*, **4**, 385 (1956).
[12] J. C. Fisher: *Transact. A.S.M.*, **47**, 451 (1955).
[13] A. H. Cottrell: *Institute of Mechanical Engineers Conference on High Rates of Strain*, (April-May 1957).
[14] W. C. Dash: *Journ. Appl. Phys.*, **27**, 1193 (1956)l.
[15] P. Haasen: *Acta Metall.*, **5**, 598 (1957).

PARTE TERZA

Semiconduttori

Properties of Semiconductors (*).

H. Y. Fan

Purdue University - Lafayette, Indiana

Introduction.

Solid materials may be divided into metals and non-metals, with respect to their conduction properties. The non-metals include insulators and semiconductors. It will be clear from a discussion of the theory of solids that, with respect to electronic conduction, there is a basic difference between metals and non-metals, whereas the division between insulators and semiconductors is arbitrary.

We shall discuss in this series of lectures the bulk properties of uniform single crystals of semiconductors. Surface phenomena and properties associated with inhomogeneous regions, such as p-n junctions, are outside the proposed scope. Several review papers on semiconductors have appeared in recent years (*) which include extensive references. We shall therefore give only references to the very recent works that are relevant to the discussion, without aiming at completeness. The discussion will be divided into the following sections: 1. Energy Band Structure, 2. Effect of Imperfections, 3. Transport Phenomena in General, 4. Transport Phenomena, 5. Infrared Optical Properties.

(*) Preparation of the paper was supported by a Signal Corps Contract.

(*) The general articles include: K. Lark-Horovitz: *Present State of Physics*, *AAAS* (1954); E. Burstein and P. H. Egli, also H. Brooks: *Advances in Electronics and Electron Physics*, (Academic Press, 1955); H. Y. Fan: *Solid State Physics*, Vol. 1 (Academic Press, 1955); O. Madelung: *Hand. d. Phys.*, XX (Berlin, 1956). Articles on specific properties and special types of semiconductors can be found in the volumes of *Solid State Physics*, edited by Seitz and Turnbull (Academic Press), and in the series *Halbleiterprobleme*, edited by W. Schottky (Braunschweig).

1. – Energy band model.

The modern theory of solids is largely based on the one electron approximation which treats individual electrons by replacing the effect of all the other electrons in the material with a suitable average potential. The problem is further simplified by considering perfect crystals and assuming that the atoms stay fixed at the regular lattice sites; crystal imperfections and atom vibrations are treated subsequently as perturbations. Thus the problem is reduced to that of an electron in a three dimensional periodic potential field. According to Bloch, the electron wave function can be written in the form

$$\psi_{\boldsymbol{k}m} = \exp[i\boldsymbol{k}\cdot\boldsymbol{r}]u_{\boldsymbol{k}m}(\boldsymbol{r}) , \tag{1.1}$$

where u is a function having the periodicity of the crystal lattice. The periodic variation of u is superimposed on the plane wave with wave number $\boldsymbol{k}$. For each value of $\boldsymbol{k}$, there is a series of different solutions for u, each indicated by a subscript m, which correspond to different energies. By varying the value of $\boldsymbol{k}$, a series of energy bands is obtained. Each of the energy bands covers a range of energy. The lower energy bands are quite narrow and are widely separated from each other; these bands accomodate the inner shell electrons of the atoms. With increasing energy, the bands broaden and may overlap each other.

It follows from the form of (1.1) that a wave function corresponding to $\boldsymbol{k}$ may be considered also as a wave function corresponding to

$$\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{g} ,$$

where $\boldsymbol{g}$ is any vector satisfying the condition that $\exp[i\boldsymbol{g}\cdot\boldsymbol{r}]$ has the periodicity of the lattice. The fact indicates that the wave function and the energy of the electron are periodic functions of $\boldsymbol{k}$. To satisfy the condition for $\boldsymbol{g}$ it is only necessary that

$$\boldsymbol{g}\cdot\boldsymbol{a}_i = 2\pi g ,$$

where $\boldsymbol{a}_i$ $(i=1, 2, 3)$ are the primitive translations of the lattice and g is an integer. This means that the vectors $\boldsymbol{g}$ are vectors of the reciprocal lattice. Since, by adding a suitable vector $\boldsymbol{g}$, any vector $\boldsymbol{k}$ can be obtained from a corresponding vector $\boldsymbol{k}'$ within the limits of

$$-\pi < \boldsymbol{k}'\cdot\boldsymbol{a}_i < \pi , \tag{1.2}$$

we need only to consider values of $\boldsymbol{k}$ bound by (2.1). The region in $\boldsymbol{k}$ space

bound by (2.1) is called the first Brillouin zone, and the vectors within this zone are referred to as reduced wave-number vectors. Since the Brillouin zone depends only on the primitive translations, it has fourteen types corresponding to the fourteen Bravais lattices. Thus the diamond-type and ZnS type structures of many important semiconductors have similar Brillouin zones which are characteristic of the face-centered cubic lattice.

The reduced wave-vectors $\boldsymbol{k}$ are discrete. The density of points, each representing a distinct $\boldsymbol{k}$, is uniform in the $\boldsymbol{k}$-space and is equal to $V/(2\pi)^3$, where V is the volume of the crystal. For each energy band, the density of electronic states in $\boldsymbol{k}$-space is

$$\varrho_{\boldsymbol{k}} = 2V/(2\pi)^3 , \tag{1.3}$$

in view of the two spin states. The volume of the first Brillouin zone in $\boldsymbol{k}$-space is $(2\pi)^3\,\Omega$, where Ω is the volume of the primitive cell of the crystal lattice. Thus the number of states in each energy band is twice the number of primitive cells in the crystal. According to the Pauli exclusion principle, each state can be occupied only by one electron. The material is a semiconductor, if the electrons just fill a number of energy bands and there is an energy gap separating the highest filled band and the empty band above it. Otherwise, the material is a metal.

The essential difference between metals and semiconductors results from the fact that the electrons in a completely filled band do not contribute to conduction. An applied field produces a current, if it changes the distribution of electrons among the various states so that there are more electrons with velocities in the direction opposite to the field. An energy band which remains filled with electrons does not contribute any change in electron distribution and hence gives no current. An electric field of ordinary magnitude is not sufficient to raise electrons across the energy gap of a semiconductor. Therefore, the conduction in an ideal semiconductor depends on the thermal excitation of electrons from the valence to the conduction band. The few electrons in the largely empty conduction band can change their states under applied field and give rise to a current; they are the conduction electrons. Also, with some vacant states or holes in the valence band, redistribution of electrons becomes possible and a current may be produced. The conduction electrons and holes are the charge carriers of a semiconductor.

The group velocity of an electron wave packet in a periodic potential field is given by

$$\boldsymbol{v} = \frac{1}{\hbar}\,\mathrm{grad}_{\boldsymbol{k}}\, E\ . \tag{1.4}$$

Under a force $\boldsymbol{F}$, due to applied electric and magnetic fields, the wave vector

of an electron changes according to

$$\frac{d\hbar \boldsymbol{k}}{dt} = \boldsymbol{F} = -e\boldsymbol{E} - \frac{e}{c}\boldsymbol{v}\times\boldsymbol{H}\,. \tag{1.5}$$

$\hbar\boldsymbol{k}$ has been called the crystal momentum. The momentum of the electron is of course $m\boldsymbol{v}$. The applied force determines the rate of change of $\hbar\boldsymbol{k}$ rather than $m\boldsymbol{v}$ because the electron is not free but is in the periodic field of the crystal. It follows from (1.4) and (1.5) that the acceleration is given by

$$\frac{d\boldsymbol{v}}{dt} = \frac{1}{\boldsymbol{m}^*}\boldsymbol{F}\,, \tag{1.6}$$

where

$$\frac{1}{\boldsymbol{m}^*} = \frac{1}{\hbar^2}\,\mathrm{grad}_{\boldsymbol{k}}\,\mathrm{grad}_{\boldsymbol{k}}\,E \tag{1.7}$$

is a tensor. $\boldsymbol{m}^*$ is called the effective mass of the electron.

Obviously, the effective mass is positive near the bottom of an energy band and negative near the top. It can be shown that the sum of $1/\boldsymbol{m}^*$ or the total acceleration of electrons in a filled band is zero. The conduction effect of an energy band with a few vacant states is the negative of the effect of the missing electrons. The missing electrons are usually from the energy levels near the top of a filled band, which are associated with negative effective masses. Their accelerations are the same as that of positively charged particles of positive effective masses. Such particles are the so-called holes. The electric current they give is the negative of the current of the missing electrons and is therefore the same as the current given by the energy band. Thus, it is possible to consider the equivalent effect of a few holes instead of all the electrons in a nearly filled band.

Theoretical calculation of the energy bands for a given crystal is a difficult problem. Even for the most investigated materials as germanium and silicon, reliable quantitative results are obtained only by combining extensive theoretical work with experimental deductions. Some information may be deduced from general symmetry considerations, i.e. the energy as function of $\boldsymbol{k}$ in the Brillouin zone has the symmetry of the reciprocal lattice. A simple deduction is that, when the maximum or minimum energy of a band does not correspond to $\boldsymbol{k}=0$, the maximum or minimum must occur for several values of $\boldsymbol{k}$ which are equivalent. Thus an energy band in a crystal of cubic symmetry may have 8 minima corresponding to equivalent $\boldsymbol{k}$ vectors, one along each of the $\langle 111\rangle$ directions; the number of minima is reduced to 4 if the $\boldsymbol{k}$ vectors end on the boundary of the Brillouin zone since each two of the 8 vectors are essentially the same reduced wave vectors. This situation has been found

for the conduction band of germanium. On the other hand, the conduction band of silicon has been found to have 6 minima corresponding to $\boldsymbol{k}$ vectors along the equivalent $\langle 100 \rangle$ directions. In general, the number of minima or maxima should be 48 for any energy band in a cubic crystal, if the $\boldsymbol{k}$ values corresponding to the minimum or maximum energy do not lie on any element of symmetry. By applying the group theory, it is possible to obtain for a given crystal the types of energy bands of various symmetries. Useful information can be obtained particularly regarding the situation where different energy bands contact each other.

In a semiconductor, the small number of conduction electrons is concentrated near the bottom of the conduction band and the holes are grouped near the top of the valence band. Therefore, it is important to know the two energy bands for only a small range of energy. Since first order terms in $\boldsymbol{k}$ are absent near a maximum or a minimum, the simplest case is where the energy for the range of interest can be approximated by a quadratic function of $\boldsymbol{k}$. With the proper choice of the co-ordinate system, we can write

$$E = E_c + ak_1^2 + bk_2^2 + ck_3^2 = E_c + \frac{\hbar^2}{2}\left(k^2/m_{1c} + k_2^2/m_{2c} + k_3^2/m_{3c}\right) \tag{1.8}$$

for the region near the bottom of the conduction band, and

$$E = E_v - (\hbar^2/2)(k_1^2/m_{1v} + k_2^2/m_{2v} + k_3^2/m_{3v}) \tag{1.8a}$$

for the region near the top of the valence band. According to (1.7), the effective mass tensor of the conduction electrons or holes has only diagonal elements, m_1, m_2, m_3, which are constants. For cubic crystals $m_1 = m_2 = m_3$.

In case the maximum or minimum energy is not at $\boldsymbol{k} = 0$ but corresponds to δ equivalent $\boldsymbol{k}_i$ vectors, k_1, k_2, k_3 are to be taken as the components of $(\boldsymbol{k} - \boldsymbol{k}_i)$ in the appropriate co-ordinate system which is different for different $\boldsymbol{k}_i$. A surface of constant energy in $\boldsymbol{k}$ space consists of δ ellipsoids as shown schematically in Fig. 1. By considering the volume between two surfaces of energies E and $(E + \mathrm{d}E)$, we get, according to (1.3) and (1.8), the num-

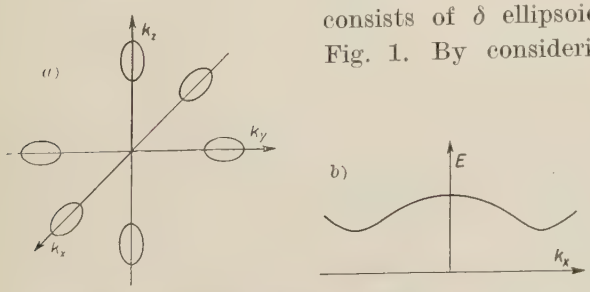


Fig. 1. – Schematic diagrams for a many-valley energy band. (*a*) Shows a constant energy surface in k space, which consists of six ellipsoids. (*b*) Shows the energy as a function of k_x.

ber of states per unit energy range per unit volume,

$$N(E) = \frac{2}{\pi^2\hbar^3} m_d^{\frac{3}{2}} \sqrt{E}\,, \tag{1.9}$$

where the parameter

$$m_d = \delta^{\frac{2}{3}}(m_1 m_2 m_3)^{\frac{1}{3}} \tag{1.10}$$

may be called the effective density-of-states mass.

In general, the energy may not be a simple quadratic function of $\boldsymbol{k}$. For example, in germanium and silicon two energy bands have the same maximum corresponding to the top of the valence band at $\boldsymbol{k} = 0$. The energy expression is

$$E = E_v - (\hbar^2/2m)\{Ak^2 \pm [Bk^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]\}.$$

Each band is doubly degenerate as usual. There would be three energy bands with the same maximum energy, except for the splitting-off of one of them due to the spin-orbit interaction. Such an energy band is shown in Fig. 2. The spin-orbit coupling is important in cases where several energy bands come together.

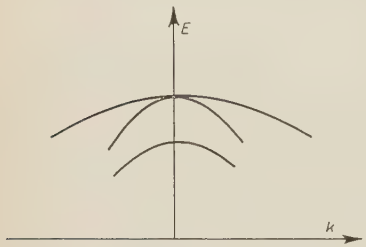


Fig. 2. – Schematic diagram of the valence band in germanium or silicon, showing the overlapping branches.

The concentration of carriers is related to the Fermi energy, ζ, which determines the electron distribution. According to the Fermi-Dirac statistics, the probability of electron occupation for an energy level E is

$$f_e = (\exp[(E - \zeta)/kT] + 1)^{-1}\,,$$

and the probability for an energy state to be vacant or occupied by a hole is

$$f_h = 1 - f_e = (\exp[(\zeta - E)/kT] + 1)^{-1}\,.$$

The concentration of conduction electrons and holes are

$$n = \int_{C.B.} f_e(E)N(E)\,\mathrm{d}E \quad \text{and} \quad p = \int_{V.B.} f_h(E)N(E)\,\mathrm{d}E, \tag{1.11}$$

respectively. When the carrier concentrations are sufficiently low that f_e and f_h can be approximated by the exponential or Maxwellian distribution functions, and when $N(E)$ can be expressed by (9), we get

$$\begin{aligned} n &= (2/h^3)(2\pi m_{dc}kT)^{\frac{3}{2}}\exp[(\zeta - E_c)/kT] = N_c \exp[(\zeta - E_c)/kT]\,, \\ p &= (2/h^3)(2m_{dv}\,kT)^{\frac{3}{2}}\exp[(E_v - \zeta)/kT] = N_v \exp[(E_v - \zeta)/kT]\,, \\ np &= N_cN_v \exp[(E_v - E_c)/2kT] = N_cN_v \exp[-E_G/kT] = n_i^2\,. \end{aligned}$$

n_i is the concentration of either electrons or holes in an intrinsic semiconductor, the carriers of which are produced by thermal excitation of electrons across the energy gap.

Experimentally, the investigations of optical properties, cyclotron resonance, and magnetoresistance are effective means for studying the structures of the energy bands. The effective mass of carriers can be obtained also from many other types of measurements such as thermoelectric power, magnetic susceptibility, specific heat, dielectric property at high frequencies, etc. The energy gap can be determined from optical absorption measurements, and from the temperature variation of intrinsic carrier concentration, as indicated by the conductivity and Hall coefficient, at sufficiently high temperatures.

2. – Lattice imperfections.

Any departure from the regular arrangement of the constituent atoms on proper lattice sites, or the presence of impurity atoms, constitutes a crystal imperfection which may bind electrons in localized orbits. A center of imperfection is a donor, if it can become more charged by releasing a bound electron. Thus a donor is either neutral or positively charged. On the other hand, a center is an acceptor if it can trap an additional electron and becomes more charged in doing so. Therefore an acceptor is either neutral or negatively charged. It follows from the electrical neutrality of the crystal as a whole that a crystal, having imperfections in positively charged donor states, should have an excess of conduction electrons over holes. Since the product of the electron and hole concentrations is a constant of the crystal for any given temperature, it means that the electron concentration will be higher than that in the intrinsic material. Under such conditions, the semiconductor is said to be extrinsic and n-type. Similarly, a semiconductor with negatively charged acceptors has more hole carriers than conduction electrons and is said to be p — type.

An imperfection center may be capable of binding several electrons in localized orbits. It is possible for a center to act as donors as well as acceptors, depending on the number of bound electrons. Furthermore, with a given number of bound electrons, the center may have various excited states. Thus the different states should be distinguished by two indices, one index giving the number of bound electrons and the other indicating the particular state. Let ε_{mn} be the total energy required to hold m electrons in the (m, n) state and let g_{mn} be the degeneracy of the state. The probability for the center to be in the (m, n) state is

$$f_{mn} = g_{mn} \exp\left[(m\zeta - \varepsilon_{mn})/kT\right]\Big\{\Big[\sum_{mn}\sum g_{mn} \exp\left[(m\zeta - \varepsilon_{mn})/kT\right]\Big\}^{-1}, \tag{2.1}$$

where ζ is the Fermi energy. Let D be the number of bound electrons that makes the center electrically neutral. Then a center carries, on the average,

$$F = \sum_m m \sum_n f_{mn} - D \,, \tag{2.2}$$

charges of $(-e)$. Let N_i be the concentration of the i-th type of imperfection centers. The condition for the electrical neutrality of the crystal,

$$n + \sum_i F_i N_i = p \,, \tag{2.3}$$

determine ζ and hence the carrier concentrations.

The situation is simplified, if with any given number of bound electrons the excited states are higher than the ground state by energies large compared to kT. Then we can approximately consider only the ground states, and eliminate in (2.1) the index n and the summation over n. If, in addition, the energy differentials, $(\varepsilon_m - \varepsilon_{m+1})$, vary by amounts large compared to kT, then f is appreciable only for those two states, j and $(j+1)$, for which $(\varepsilon_j - \varepsilon_{j-1})$ is the closest to ζ. It can be shown that approximately

$$f_j \sim \left\{1 + \frac{g_{j-1}}{g_j} \exp\left[\frac{\varepsilon_j - \varepsilon_{j-1} - \zeta}{kT}\right]\right\}^{-1}, \qquad f_j \sim 1 - f_{j-1} \,. \tag{2.4}$$

In the electron energy diagram the imperfection can be represented by energy levels $E_j = \varepsilon_j - \varepsilon_{j-1}$ as shown in Fig. 3. E_j is the energy required to trap the j-th electron.

As an example, we consider imperfection centers which can only trap one electron per center. Each center is then represented by a single level, $j = 1$. In the expression (2.2), the value of D is unity if the centers are donors and zero if the centers are acceptors. Consider the case of donors. (2.2) and (2.4) give

$$F = f_1 - 1 = -\left(1 + g_D \exp\left[\frac{\zeta - E_D}{kT}\right]\right)^{-1}, \tag{2.5}$$

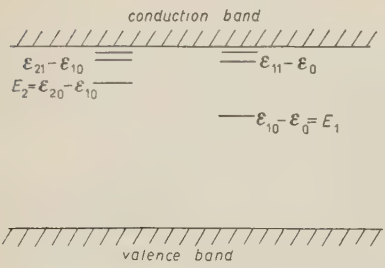


Fig. 3. – Schematic energy level diagram of a semiconductor, showing the energy levels of an imperfection center.

where $g_D = g_1/g_0$ and E_D stands for $(\varepsilon_1 - \varepsilon_0)$, the energy required to hold the electron to the donor. Let N_D be the concentration of donors and let there be only one kind of imperfection. Then equa-

tion (2.3) becomes

$$n - p = N_D \left(1 + g_D \exp\left[\frac{\zeta - E_D}{kT}\right]\right)^{-1} . \tag{2.6}$$

With sufficiently large N_D, the semiconductor becomes n-type with $n \gg p$. By neglecting p and using (1.12), we get the usual expression for the carrier concentration,

$$\frac{n^2}{N_D - n} = \frac{N_c}{g_D} \exp\left[-\left(\frac{E_c - E_D}{kT}\right)\right], \tag{2.7}$$

$(E_c - E_D)$ is the energy required to remove the bound electron from the donor to the conduction band. It is often called the ionization energy of the donor. If the excited states of the donor are taken into account, then we get

$$F = \sum_n f_{1n} - 1 = \sum_n g_{1n} \exp\left[(\zeta - \varepsilon_{1n})/kT\right] \cdot \cdot \left(\sum_n g_{1n} \exp\left[(\zeta - \varepsilon_{1n})/kT\right] + 1\right)^{-1} - 1 . \tag{2.8}$$

This equation, written in the form of (2.5), will give a g_D and an E_D which vary with the temperature.

Theoretical treatment of the states of imperfection centers is, in general, very difficult. Only one case has been treated fairly satisfactorily, i.e. a substitutional impurity atom which differs by one valence from the regular atom. Consider an impurity atom with one more valence electron. Without the electron, the valence band can be just filled. The extra electron has to go into the conduction band but may also be bound in the neighborhood of the impurity atom which has an extra nuclear charge. An average attractive potential produced by the extra nuclear charge in the crystal is

$$V(r) = - e^2/Kr , \tag{2.9}$$

where K is the dielectric constant. Thus, the extra electron may be bound in hydrogenic states with a Bohr radius,

$$r_0 = \frac{m}{m^*} Ka , \tag{2.10}$$

where a is the Bohr radius of hydrogen atom and m^* is the effective mass of conduction electrons. Most semiconductors have relatively large dielectric constants and effective masses comparable or smaller than m. Thus r_0 may be several times the interatomic distance in the crystal. It is this circum-

stance that justified the expression (2.9) for the potential, since the dielectric constant is an average property and can be used only when the electron orbit covers many cells. The ionization energy, the energy required to release the electron to the conduction band is

$$E_c - E_D = E_H \frac{m^*}{m} \frac{1}{K^2}, \tag{2.11}$$

where E_H is the ionization energy of hydrogen.

In the case of a substitutional foreign atom with one less valence electron, there will be a vacant state or hole in the valence band. The smaller nuclear charge on the impurity atom gives an attractive potential for positive charge and tends to bind the hole. Expressions (2.10) and (2.11) apply also in this case, and the effective mass of the holes in the valence band should be used. The ionization energy of the hole is the energy required to bind an electron from the valence band. In the electron energy diagram, the acceptor is represented by a level E_A. Expression (2.11) gives $(E_A - E_v)$.

The simple consideration is justified by more rigorous treatment which shows that the wave function of the impurity state can be expressed as the wave function for the lowest state of the corresponding energy band, limited by an envelope. The envelope function is determined by a Schrödinger-type equation with the effective mass of the carrier replacing the free electron mass. When the energy band has spherical surfaces of constant energy in $\boldsymbol{k}$ space, the equation is like the wave equation for the hydrogen atom, and results are given just by (2.10) and (2.11). More complicated solutions have been obtained for cases of more complicated energy bands such as the conduction and valence bands in silicon and germanium.

The treatment based on (2.9) gives satisfactory explanation of small ionization energies observed in semiconductors with large dielectric constants, particularly for impurities of group III and group V elements in silicon and germanium. However such a simple model gives impurity states which depend only on the energy band structure but not on the impurity atom. Actually, even in silicon and germanium, the ionization energies of impurity elements from the adjacent columns vary appreciably. This is not surprising in view of the fact that (2.9) can not be expected to apply within the cell occupied by the impurity atom, but it is not easy to estimate the potential theoretically.

There is yet no reliable method for estimating the energy levels of other types of imperfections. Some qualitative understanding of the experimental observations may be gained by reasonable arguments. For example, Zn, Fe, Ni, Co, and Mn each gives two acceptor levels in germanium, while Cu and Au each gives three acceptor levels [1]. It is assumed that all these impurities are substitutional and that the acceptor action is determined by the tendency to

complete the tetrahedral valence bonds with four nearest germanium atoms. Thus the atoms with a S^2 configuration may act as double acceptors and those with a S^1 configuration may be triple acceptors. However, other arguments have to be invoked to explain the fact that Au gives also a fourth level in germanium, which is a donor level; furthermore two donor levels but no acceptor levels have been observed for Au in silicon.

As to lattice defects, the simplest type is isolated vacancies in the lattice sites and atoms in interstitial positions. It is well known that a negative ion vacancy in alkali halides forms F centers by trapping an electron and becoming electrically neutral, it is therefore a donor. A positive ion vacancy becomes neutral by trapping a hole to form a V_1 center, it is therefore an acceptor. A direct method for producing Frenkel defects is the irradiation with high energy electrons or nucleons. The particles displace atoms from the lattice sites by collision. Donor as well as acceptor levels are produced by irradiations in germanium, silicon and other semiconductors. It is believed that the donor levels are associated with interstitials while the acceptor levels belong to the vacancies. In valence semiconductors, dislocations may be expected to produce acceptor levels, since an edge dislocation gives a row of atoms with « dangling bonds». Two acceptor levels are produced in germanium by plastic deformation; at least one of them is associated with the introduced dislocations.

Experimentally, imperfection states can be investigated by making various types of observations. Within suitable temperature ranges, the variation of carrier concentration is controlled by the ionization energy of imperfections, according to some relation such as illustrated by (2.7) for the simplest case. Photo-ionization and excitation may be revealed in the infrared absorption spectrum of the crystal; optical ionization energy may also be obtained from the threshold of the photoconductivity which is associated with the presence of imperfections. Important information may be obtained from spin paramagnetic resonance of the bound electrons, e.g., when the impurity atom has a nuclear spin, hyperfine splitting of the resonance line occurs, and the magnitude of the splitting gives an estimate of the electron wave function near the nucleus. Polarizability of imperfection centers may be obtained sometimes by measuring the dielectric constant of the crystal at suitable frequencies. The charge carried by each imperfection center may be estimated from the effect of a known concentration of centers on the carrier mobility, since Coulomb scattering is proportional to the square of the charge; the fact that some impurity atoms act as multiple acceptors in germanium has been demonstrated by studying the effect of a given impurity concentration in samples having different Fermi levels.

Finally, imperfection centers are important for the electrical properties of semiconductors under non-equilibrium conditions. Usually, the hole-electron recombination rate in most samples of any semiconductor is determined either

by some impurity or by some type of lattice defects which can readily capture electrons from the conduction band and holes from the valence band. Imperfection centers which capture one type of carriers more easily give rise to trapping effect, i.e., builds up a fixed space charge which decays at much slower rate than the recombination rate of free carriers. The capture cross-section for carriers can be determined by studying phenomena associated with non-equilibrium carrier concentrations such as injection phenomena and photoconductivity.

3. – Transport phenomena, general.

3 1. *Boltzmann equation.* – In a perfectly periodic potential field, carriers move freely with velocity (1.4) and are accelerated freely according to (1.5). Resistance to the flow of carriers and the energy loss associated with it arise from the interactions or collisions of the carriers with the lattice, which are produced by disturbances in the periodic potential. Such disturbances result from lattice vibration and imperfections. The transport properties of carriers can be calculated once the carrier distribution among the various electronic states is known. The Boltzmann equation for the distribution, well known in the kinetic theory of gases, takes the form

$$\frac{\partial}{\partial t} f(\boldsymbol{k}, \boldsymbol{r}) = -\frac{q}{\hbar}\left(\boldsymbol{E} + \frac{1}{c\hbar}\operatorname{grad}_k E \;\; \boldsymbol{H}\right) \cdot \operatorname{grad}_k f - \frac{1}{\hbar}\operatorname{grad}_k E \cdot \operatorname{grad}_r f + \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}, \tag{3.1}$$

where q is the charge of the carrier. The first term on the left hand side gives the effect of applied electric and magnetic fields. The effect of any temperature variation is taken into account by the second term. The third term gives the effect of collisions. It is usually assumed that the effect of collisions can be expressed in the form

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{f - f_0}{\tau}. \tag{3.2}$$

where f_0 is the distribution function under equilibrium conditions. τ is called the relaxation time and may be a function of $\boldsymbol{k}$. The assumption, known as the existance of a relaxation time, is not always valid, but the problem becomes prohibitively complicated without the assumption.

Ordinarily, applied fields and temperature gradients are such that the deviation of distribution function from f_0 is small. Then f can be replaced by f_0 in the second term and in the term containing the electric field. For a

uniform crystal, $\mathrm{grad}_r f$ in the second term can be written as

$$\mathrm{grad}_r f_0 = \frac{\partial f_0}{\partial E} T \frac{\partial}{\partial T}\left(\frac{E-\zeta}{T}\right) \mathrm{grad}_r T \,. \tag{3.3}$$

The term for the magnetic field vanishes for $f - f_0$, since, with f_0 depending only on E, the vector $\mathrm{grad}_k f_0$ has the same direction as the vector $\mathrm{grad}_k E$. Therefore the effect of the magnetic field is associated only with the departure from f_0. Let

$$f = f_0 - \frac{\partial f_0}{\partial E} g(\boldsymbol{k}) \,. \tag{3.4}$$

If the applied electric and magnetic fields and the temperature variation in the crystal are maintained constant, then $\partial f/\partial t = 0$ and we get from the Boltzmann equation

$$g(\boldsymbol{k}) = \frac{\tau}{\hbar} \mathrm{grad}_k E \cdot \left[q\boldsymbol{E} + T \,\mathrm{grad}_r\left(\frac{E-\zeta}{T}\right)\right] + \frac{q\tau}{c\hbar^2} (\mathrm{grad}_k E \times \boldsymbol{H}) \cdot \mathrm{grad}_k g \,. \tag{3.5}$$

This gives $g(\boldsymbol{k})$ directly if there is no magnetic field. In the presence of a magnetic field, the differential equation has to be solved. Solutions have been obtained for the cases: 1) scalar effective mass, i.e., spherical surfaces of constant energy in $\boldsymbol{k}$ space, and 2) ellipsoids of constant energy in $\boldsymbol{k}$ space and τ dependent only on energy. In more general cases, solutions in the form of a series in powers of H or $1/H$ have to be used.

When an ac electric field $\boldsymbol{E} = \boldsymbol{E}_0 \exp[i\omega t]$ is applied, with or without a constant magnetic field, we can write for steady state conditions

$$g(\boldsymbol{k}) = g_0(\boldsymbol{k}) \exp[i\omega t] \quad \text{and} \quad \frac{\partial f}{\partial t} = -i\omega \frac{\partial f_0}{\partial E} g(\boldsymbol{k}) \,. \tag{3.6}$$

It is easily seen that $g_0(\boldsymbol{k})$ can be obtained from the solution for dc field E_0 by replacing τ with $\tau/(1+i\omega\tau)$. This applies for example, to the analysis of cyclotron resonance measurements.

All properties associated with the transport of carriers can be calculated once f or g is obtained. The relaxation time, τ, must be known. We shall discuss the properties of τ for different mechanisms of scattering.

3·2. *Lattice scattering.* — Lattice vibration is a major mechanism of scattering. Carrier scattering is accompanied by the absorption or emission of a phonon, a quantum of vibrational energy. The transition probability obeys the selection rule

$$\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{q} \,, \tag{3.7}$$

where $\boldsymbol{q}$ is the wave number of the phonon, and the plus and minus signs apply respectively to phonon absorption and phonon emission processes. Furthermore, the conservation of energy requires

$$E(\boldsymbol{k}') = E(\boldsymbol{k}) + \hbar\omega_{\boldsymbol{q}}\,, \tag{3.8}$$

$\hbar\omega_{\boldsymbol{q}}$ being the phonon energy.

The normal modes of lattice vibration have always three acoustical branches with frequencies ranging from zero up. The scattering due to the acoustical modes are important for homopolar semiconductors as for metals. In semiconductors, this type of scattering gives

$$\frac{1}{\tau} = \frac{4\sqrt{2}}{9\pi}\frac{C^2kTm_d^{\frac{3}{2}}}{\hbar^4u^2d}E^{\frac{1}{2}}\,, \tag{3.9}$$

where u is the sound velocity, d is the density of the crystal, $m_d = (m_1m_2m_3)^{\frac{1}{3}}$ in case of tensor effective mass, and E is the carrier energy measured from the edge of the energy band. The parameter C depends on the interaction potential which is difficult to calculate. It can be related to other properties of the crystal by the deformation theory of Bardeen and Shockley.

For acoustical phonons $\hbar\omega_{\boldsymbol{q}} = hu\boldsymbol{q}$. With $E(\boldsymbol{k}) = \hbar^2k^2/2m^*$, conditions (3.7) and (3.8) limit $\boldsymbol{q}$ to be of the order of k. The energy E of carriers at ordinary temperatures is of the order of kT for which k and hence q correspond to wave lengths much larger than the lattice spacing. The perturbation potential produced at any point by such long waves is therefore nearly the same as that produced by a homogeneous elastic strain equal to the local strain. For a cubic crystal subject to a small homogeneous strain, the edge of an energy band, if situated at the center of the Brillouin zone, can be expressed in the form

$$E_B = E_{B_0} + E_1\Delta\,, \tag{3.10}$$

where Δ is the lattice dilatation. E_1 is called the deformation potential. $E_1\Delta$ may be taken approximately as the effect of strain on all energy levels close to the band edge, and the interaction potential for long acoustical waves can be then given in terms of E_1. In fact it turns out that the parameter C is simply

$$C = \frac{3}{2}E_1\,. \tag{3.11}$$

The linear dependence of $1/\tau$ on temperature is obtained also because of the long wavelength of the phonons involved. The number of phonons with an energy $\hbar\omega_{\boldsymbol{q}} \ll kT$ is proportional to kT.

In the case where the edge of the energy band is not at the center of the Brillouin zone, the shift of E_B is no longer given by the simple expression (3.10) but depends on various strain components. In a cubic crystal, two parameters are required when the band edge is situated on ⟨100⟩ or ⟨111⟩ axes, as in the conduction band of silicon or germanium [2]. One of the parameters, Σ_d represents the shift due to dilatation in the two directions normal to the axis while the other, Σ_u, represents that due to a stretch along the axis and a contraction in the normal directions. Two relaxation times have to be used, for the directions parallel and perpendicular to the axis, respectively. Both are proportional to $E^{\frac{1}{2}}T$. Since a shear strain shifts the band edge by different amounts at the different points in $\boldsymbol{k}$ space, the conductivity of the sample may change and become anisotropic. This effect, known as piezoresistance [3], is intimately related to Σ_d and Σ_u and has been observed in n-type germanium and silicon. The measurement of this effect can be used in conjunction with τ or mobility for the determination of the coefficients.

Although the vibrational modes effective for scattering are normally those of small $\boldsymbol{k}$, in case of a multi-valley energy band the scattering of carriers from one valley to the other requires phonons of relatively large $\boldsymbol{k}$ which should be roughly equal to the separation between the centers of different valleys in $\boldsymbol{k}$ space. In this case, the interaction potential cannot be obtained from the deformation potential theory.

Crystals with more than one atom per primitive cell have optical branches of vibrational modes. In these modes of vibration, the atoms in the same cell vibrate with different phases. These modes of vibration may also contribute to intervalley scattering in the case of a multi-valley energy band. The electron mobility given by lattice scattering in silicon is found to be proportional to $T^{-2.6}$, in contrast to the $(E^{\frac{1}{2}}T)^{-1}$ or $T^{-\frac{3}{2}}$ dependence of τ expected for intravalley scattering by acoustical vibrations. Intervalley scattering may be the explanation. For ionic crystals, the optical modes are the most important mechanism of scattering, since the polarization produced gives a strong interaction with the carriers; the expression of τ valid for $T \ll \hbar\omega/k$ is [4]

$$\frac{1}{\tau} = \left(\frac{1}{K_\infty} - \frac{1}{K_0}\right)\frac{e^2}{\hbar^{\frac{3}{2}}}(2m\omega)^{\frac{1}{2}}\left(\exp\left[\frac{\hbar\omega}{kT}\right] - 1\right), \tag{3.12}$$

where K_0 and K_∞ are the static and high frequency dielectric constants respectively, and $\hbar\omega$ is the energy of long wavelength optical phonons. When the crystal is partially ionic in character, this expression can be used with an effective charge. In homopolar crystals, the effect of optical modes is difficult to estimate and its importance is uncertain.

The valence bands in germanium and silicon present a more complicated

case. The two branches of the energy band, degenerate at $\boldsymbol{k} = 0$, give two types of holes. There are scattering within each branch as well as between the two branches. The problem has been treated recently by EHRENREICH and OVERHAUSER [5] taking into account both acoustical and optical phonons. The results involve two adjustable parameters.

3 3. *Scattering by charged centers.* – The scattering effect of ionized impurity atoms is usually treated as that of a charged center in a medium having the dielectric constant of the crystal, since the long range Coulomb interaction should be more effective than any potential perturbation local to the atom. The model may be also applicable to charged defect centers in the lattice. The scattering problem has been solved classically by CONWELL and WEISSKOPF, using simple Coulomb potential. The well known divergence of total cross-section for Rutherford scattering was avoided by limiting the scattering effect of each center to a sphere, the radius of which was taken to be one half of the average distance between impurity atoms. BROOKS and HERRING treated the problem using the Born approximation, and introduced a shielding factor in the potential to take into account the screening effect of carriers. The result is

$$\frac{1}{\tau} = \frac{\pi\hbar}{m^{*\frac{1}{2}}} \frac{1}{(2E)^{\frac{3}{2}}} \frac{e^4 N}{K^2} \left[\ln (1+b) - \frac{b}{1+b} \right], \tag{3.15}$$

where

$$b = \frac{8\pi m^* K k T}{e^2 \hbar^2 n} E .$$

n is the carrier concentration, and N is concentration of charged centers. The expression is derived for spherical energy surfaces. The scattering is highly anisotropic, being strong for small angles of scattering. For this reason, $1/\tau$ no longer depends simply on E in case of ellipsoidal energy bands. A more general expression for $1/\tau$ has not yet been derived. The Born approximation becomes poor for small carrier energies, comparable or smaller than the ionization energy of the impurity atom. Recently, more rigorous treatments have been attempted which are necessary for low temperature applications [6].

3 4. *Other types of scattering.* – Scattering by neutral impurities has been treated by ERGINSOY using the model of hydrogen atom in a dielectric medium. The experimental data for the scattering of low energy electrons by hydrogen atom was used to obtain an approximate expression for the cross-section. The relaxation time obtained is independent of the energy of carriers. Recent work of SLAR [7] indicates that some energy dependence may be expected. This type of scattering should be important at low temperatures when the

lattice scattering becomes small and the effective impurities become unionized. Scattering by dislocations, as estimated from the deformation potential associated with the strain, is small but can be much larger if the dislocations become electrically charged by virtue of the localized electronic states that they introduce; these predictions are in qualitative agreement with experimental observations made on plastically deformed germanium. In the intrinsic temperature range where both electron and hole carriers are present in large concentrations, the mutual scattering of electrons and holes which drift in opposite directions may become appreciable. This type of scattering may be estimated in a similar way as the scattering by ionized centers. In the presence of several types of scattering, $1/\tau$ is the sum of the contributions of each type

$$1/\tau = \sum_j 1/\tau_j \,. \tag{3.14}$$

4. – Transport phenomena.

4·1. *Conductivity and Hall coefficient.* – Electrical current arises as the carriers acquire, in addition to the random thermal velocity, a drift velocity under the force exerted by the applied field. The conductivity tensor expresses the linear relationship between the current density and the electrical field

$$j_i = \sum_j \sigma_{ij} E_j \,. \tag{4.1}$$

The mobility tensor, defined by

$$\mu = \sigma/en \,, \tag{4.2}$$

expresses the relationship between the electric field and the « average » drift velocity of the carriers. The theoretical expression for current density is

$$\boldsymbol{j} = \frac{2}{(2\pi)^3} \int q\boldsymbol{v} f(\boldsymbol{k}) \, \mathrm{d}\Omega_k , \tag{4.3}$$

and the conductivity tensor is obtained by substituting for f the solution of Boltzmann equation for the case of an applied electric field alone.

For cubic crystals, the conductivity is a scalar by symmetry. The expression of conductivity becomes

$$\sigma = -\frac{e^2}{3} \cdot \frac{2}{(2\pi)^3} \int \tau v^2 \frac{\partial f_0}{\partial E} \, \mathrm{d}\Omega_k \,. \tag{4.4}$$

If the carriers are strongly degenerate, the mobility expression can be written as

$$\mu = \frac{e}{3n} (\tau v^2)_\zeta \, g(\zeta) \,, \tag{4.5}$$

where $(\tau v^2)_\zeta$ is the value of τv^2 averaged over the Fermi surface, and $g(\zeta)$ is the density of states at the Fermi surface. For non-degenerate carriers, we get

$$\mu = \frac{e}{3kT} \langle \tau v^2 \rangle \,. \tag{4.6}$$

For the simple case, where the constant energy surfaces are spherical and τ depends merely on energy, the expressions reduce to

$$\mu = \frac{e\tau_\zeta}{m^*} \qquad \text{and} \qquad \mu = \frac{e \langle \tau v^2 \rangle}{m^* \langle v^2 \rangle} \,, \tag{4.7}$$

for the degenerate and non-degenerate carriers, respectively.

To determine experimentally the two quantities, carrier concentration and mobility, another independent measurement is required besides the conductivity. There is available a method, first employed by Shockley and Haynes, for measuring the drift mobility of the minority carriers. In order to obtain the mobility of carriers in a crystal, it is necessary to use another crystal of the opposite type, making sure that the scattering is the same in the two crystals. Usually, the conductivity measurement is supplemented with measurement of the Hall coefficient which gives an estimate of the carrier concentration.

Hall effect is the existence of an electric field component perpendicular to the current in the presence of a magnetic field normal to the plane of the current and the electric field. The transverse electric field, required to balance the Lorentz force, is proportional to the magnetic field and the drift velocity of the carriers. The Hall coefficient, R, is defined as the ratio of the transverse electric field to the product of the magnetic field and current density, therefore it is inversely proportional to the carrier concentration. For the case of spherical energy surfaces,

$$R = \frac{B}{A^2 + B^2} \frac{1}{H} \,, \tag{4.8}$$

where

$$A = -\frac{e^2}{3} \frac{2}{(2\pi)^3} \int \frac{v^2 \tau}{1 + (\delta\tau)^2} \frac{\partial f_0}{\partial E} \, \mathrm{d}\Omega_k \,,$$

$$B = -\frac{e^2}{3} \frac{2}{(2\pi)^3} \delta \int \frac{v^2 \tau^2}{1 + (\delta\tau)^2} \frac{\partial f_0}{\partial E} \, \mathrm{d}\Omega_k \,,$$

$$\delta = \frac{qH}{m^* c} \,.$$

The expression for R can be written in the form

$$R = \frac{\mu_H}{\mu}\frac{1}{cqn}, \tag{4.9}$$

μ_H is known as the Hall mobility. For degenerate carriers, $\mu_H = \mu$. For classical distribution of carriers, simple results are obtained for the limiting cases:

$$\mu_H/\mu = 1 \qquad \text{for} \quad \delta\tau \gg 1. \tag{4.11}$$

$$\mu_H/\mu = \frac{\langle v^2\tau^2\rangle}{\langle v^2\rangle} \qquad \text{for} \quad \delta\tau \ll 1. \tag{4.12}$$

In case there are carriers in several energy bands, the expression for R is

$$R = \frac{\sum B}{(\sum A)^2 + (\sum B)^2}\frac{1}{H}. \tag{4.13}$$

It may happen that the magnetic field gives a limiting case for certain carriers but not for others. This effect is responsible for some unusual variation of Hall coefficient with magnetic field observed in p-type germanium which has two types of holes with quite different effective masses.

As an illustration of more complicated cases, a many-valley energy band in cubic crystals has

$$\frac{\mu_H}{\mu} = 3\langle E\rangle\left(\frac{2\langle E\tau_\parallel\tau_\perp\rangle}{m_\parallel m_\perp} + \frac{\langle E\tau^2\rangle}{m^2}\right)\left(\frac{\langle E\tau_\parallel\rangle}{m_\parallel} + \frac{2\langle E\tau_\perp\rangle}{m_\perp}\right)^{-2}, \tag{4.14}$$

where $m_\parallel$ and $m_\perp$ are the components of mass tensor parallel and perpendicular to the axis of spheroidal energy surface in each valley. It is assumed that the scattering can be characterized by two relaxation times $\tau_\parallel(E)$ and $\tau_\perp(E)$ for carriers in each valley.

In the limit of very high H, the value of μ_H/μ always approaches unity [8]. Otherwise, the value depends on the energy band structure as well as the relaxation time. Its variation presents a difficulty for the determination of carrier concentration from the Hall effect. The temperature dependence of intrinsic carrier concentration is controlled by the exponential factor $\exp[-E_g/2kT]$. In the temperature range where the semiconductor is extrinsic, the carrier concentration may be approximately constant at high temperatures, exhaustion range, and varies as $\exp[-E_i/2kT]$ or $\exp[-E_i/kT]$ at sufficiently low temperatures, E_i being the ionization energy of the impurity. Therefore, experimental data are usually presented in the form $\log R$

and (log ϱ) versus $1/T$. Fig. 4 shows a sketch of such curves. The slope of log R curve at the high temperatures, (1), corresponds to $E_a/2k$ approximately. The exhaustion range, (2), revealed by a constant portion of the log R curve gives the concentration of effective impurities. The impurity ionization energy can be estimated from the slope at low temperatures, (3).

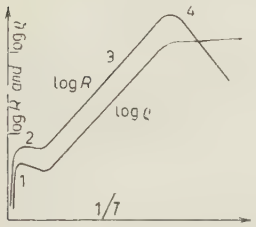


Fig. 4. – General behaviors of Hall coefficient, R, and resistivity, ρ, of a semiconductor as functions of temperature.

At sufficiently low temperatures, (4), the resistivity stops to increase according to the ionization energy of the impurity and varies much more slowly with temperature. Meanwhile, the Hall coefficient shows a maximum and drops with further cooling. These phenomena were first observed in germanium by HUNG and were attributed by him to conduction by electrons or holes in impurity states. Similar phenomena have since been observed in many other semiconductors. Theoretically, conduction in the states of randomly distributed impurity atoms presents a difficult problem which has not yet been solved satisfactorily [9]. A theory is presented by MOTT in an article of the present series.

4·2. *Magnetoresistance.* – An applied magnetic field not only gives rise to an electric field component transverse to the current but also changes for a given current the parallel component of the electrical field. The effect, known as magnetoresistance, arises from a distribution of drift velocity. The transverse electric field compensates on the average the Lorentz force, but the compensation is not complete for individual carriers of different drift velocities. The resulting lateral deflection of the path increases the resistance. The lateral deflection is proportional to $\tau H/m^*$ and the effect of a small deviation on the drift is a second order effect. Therefore, the magnetoresistance effect is expected to depend on $\mu^2 H^2$ at small magnetic fields.

In the simple case of a scalar effective mass and isotropic τ, i.e. τ varies only with energy, the magnetoresistance effect vanishes for strongly degenerate carriers, since the drift mobility depends only on one relaxation time, τ_2. For carriers having classical distribution, the conductivity under transverse magnetic field is given by the expression

$$\sigma_H = \frac{j_H^2}{\boldsymbol{j}_H \cdot \boldsymbol{E}} = \frac{A^2 + B^2}{A}, \tag{4.15}$$

where A and B are as defined in (4.8). It differs from σ_0 if τ varies with energy. The difference $(\sigma_0 - \sigma_H)$ is second order in H at small magnetic fields

and approaches a constant value at large fields. There is no longitudinal effect, i.e. σ is not changed by a magnetic field parallel to the current.

Since the magnetoresistance effect comes from the variation of τ/m^* among carriers, it is sensitive to the character of the energy band structure and the collision processes, in particular, a deviation of either one from spherical symmetry gives rise to a longitudinal effect. The magnetoresistance effect is given in terms of $(\Delta\varrho/\varrho_0 H^2)$ for different directions of j and H. More generally, the effect can be expressed by a magnetoconductivity tensor, defined by

$$\sigma_{\alpha\beta\gamma\delta} \equiv \frac{1}{2}\frac{\partial^3 j_\alpha}{\partial E_\beta \partial H_\gamma \partial H_\delta}. \tag{4.16}$$

One of the first evidences of non-spherical conduction band in germanium was obtained from magnetoresistance investigations. The multi-valleyed structures of conduction band in germanium and silicon can be deduced from the observed variations of $\Delta\varrho$ with the current and field directions [10]. The behaviour of magnetoresistance in n-type silicon is shown in Fig. 5. A curve for p-type silicon is also shown for comparison.

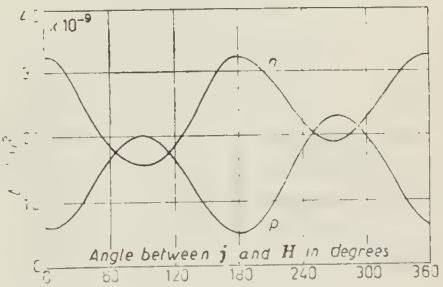


Fig. 5. – Magnetoresistances of n- and p-type silicon as functions of the direction of the magnetic field, with current in the [111] direction. After PEARSON and HERRING: *Physica*, **20**, 975 (1954).

The variation of magnetoconductivity with field strength can be calculated by a method [11] applicable for arbitrary energy surfaces, provided the relaxation time depends only on the energy. The conductivity approaches a saturation limit at high fields [12]. However, the effect of orbital quantization sets in at fields $H > m^*c/e\tau$ when the cyclotron frequency eH/m^*c becomes comparable or larger than the collision frequency $1/\tau$. The relaxation time is affected as the carriers start to execute circular motion. This effect prevents the saturation of the conductivity [13], and for certain types of scattering mechanism a negative magnetoresistance may be expected. If the carriers are degenerate, the magnetoconductivity may oscillate with varying field as the Fermi level crosses the quantized energy levels. This effect is similar to the de Haas-van Alphen effect in magnetic susceptibility, and has been reported recently for n-type InSb which has a small electron effective mass [14]. In addition, the quantum effect may change the impurity ionization energy [15] and split degenerate branches of an energy band which have different effective masses [16].

4 3. *Thermoelectric power.* – Thermoelectric power is one of the important properties to investigate. It indicates the type of conduction, n- or p-type, and yields an estimate of the density-of-states mass. The effect arises from the tendency of the carriers to diffuse from the hotter end of a material to the colder end. When a temperature gradient is maintained in a specimen, an electrical field must be built up if there is to be no current. In an isotopic material or a cubic crystal, the electrical field is along the direction or opposite to the direction of the temperature gradient, depending upon whether the carriers are electrons or holes.

It can be shown that the thermoelectric power of a thermocouple consisting of two materials A and B is given by

$$Q_{AB} = \frac{1}{q}\left[\left(\frac{K_2 - K_1\zeta}{K_1 T}\right)_A - \left(\frac{K_2 - K_1\zeta}{K_1 T}\right)_B\right], \tag{4.17}$$

where

$$K_m = \frac{1}{4\pi^3}\int \tau v_x^2 E^{m-1}\frac{\partial f_0}{\partial E}\,\mathrm{d}\Omega_k\,. \tag{4.18}$$

Thermodynamic considerations prescribe a relation between Q_{AB} and the Thomson coefficients of the individual materials:

$$\frac{\mathrm{d}Q_{AB}}{\mathrm{d}T} = \frac{\mu_A - \mu_B}{T} \tag{4.19}$$

suggesting that Q_{AB} may be considered as the difference between two quantities, each one of which is the property of one material and is referred to as the absolute thermoelectric power of that material. It follows from (4.7) that the expression for the absolute thermoelectric power may be written

$$qTQ = (K_2/K_1) - \zeta\,. \tag{4.20}$$

K_2/K_1 can be calculated if the dependence of τ on energy is known. The measurement of Q will then yield the value of ζ, from which the density-of-state mass can be obtained. For strongly degenerate carriers, $K_2/K_1 = \zeta$ in the first approximation, therefore Q is small. For carriers obeying classical statistics, $\zeta < 0$, and

$$K_2/K_1 = \langle \tau v^2 E\rangle / \langle \tau v^2\rangle \tag{4.21}$$

is small compared to $|\zeta|$ if the carrier concentration is sufficiently low.

The above discussion overlooks an important effect, i.e. a temperature

variation affects not only the carriers but also the lattice vibration. Phonons as well as carriers tend to diffuse from hotter to colder regions. The phonons exert, through collisions, a drag on the carriers, thus enhancing the thermoelectric power [17]. The importance of this effect was first established by the extension of thermoelectric measurements on germanium to low temperatures. More recently, this has been confirmed for a number of semiconductors including silicon and ZnO. The thermoelectric power consists of two terms, Q_e and Q_p, which are associated with pure electronic and phonon-drag effects respectively. At low temperatures, the observed Q is often many times larger than the calculated Q_e as shown in Fig. 6 for a germanium sample. The difference can be largely accounted for by Q_p. The phonon-drag effect depends on the electron-phonon interaction and the relaxation time of phonons. The interaction is effective for low frequency phonons which have wavelengths comparable to the electron wavelength. The relaxation time of these phonons, as determined by their collisions with other phonons, increases rapidly with decreasing temperature. The fact explains the increased importance of phonon-drag effect at low temperatures. Phonons are also scattered at the boundaries of the specimen and Q will show a dependence on the sample dimension if the dimensions are comparable to the phonon mean free path. Such size effect has been observed as expected and provides a clear demonstration of the phonon effect. The effect of non-equilibrium phonon distribution has been found to be important also for the Nernst effect of germanium.

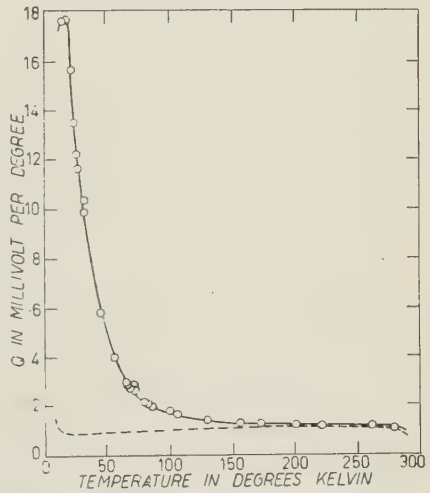


Fig. 6. – Thermoelectric power of a p-type germanium filament. The dashed curve is calculated Q_e. Measurement by GEBALLE, given in the paper by C. HERRING: *International Conference on Semiconductors and Phosphors*, Garmisch, 1956 (in print).

5. – Optical properties.

5 1. *Introduction.* – Semiconductors show interesting optical behaviors, the investigation of which yields a variety of important information [18]. The mechanisms determining the optical properties may be classified in the following way.

1) Intrinsic excitation. Excitation of electrons from the valence to the conduction band gives rise to strong absorption which is responsible for the metallic reflection shown by many semiconductors in the optical region. The absorption has a long wavelength threshold, the intrinsic absorption edge, corresponding to a photon energy equal or larger than the energy gap. At longer wavelengths, the polarization associated with the electrons in valence band approaches a constant value. For crystals of low impurity concentration, the refractive index and dielectric constant have throughout the infrared region constant values which are determined largely by the susceptibility associated with the valence band electrons.

2) Carrier effect. Carriers give rise to absorption and polarization which are related to the low frequency electrical conduction. Usually the effect becomes pronounced outside of the intrinsic absorption threshold and increases with wavelength. In addition, when there are several branches in the energy band, carrier excitation between the different branches may be observed.

3) Excitation of imperfection centers. Carriers which are localized on impurities or lattice defects give rise to absorption associated with excitations or ionization.

4) Lattice vibration. Beside the electronic processes, absorption bands may be produced by lattice vibrations. The absorption is strong when the crystal is to some extent ionic in character. Impurities and lattice defects may give rise to additional absorption bands.

5) Inhomogeneity. We include in this category effects such as that produced by mechanical strains associated with lattice defects and by aggregates of impurity atoms.

5·2. *Intrinsic excitation.* – Intrinsic excitation obeys the selection rule

$$\boldsymbol{k}_1 = \boldsymbol{k}_0 + \boldsymbol{k}_\lambda \sim \boldsymbol{k}_0 \,, \tag{5.1}$$

where $\boldsymbol{k}_0$ is the wave number of initial state in the valence band, $\boldsymbol{k}_1$ is that of the final state in the conduction band, and $\boldsymbol{k}_\lambda$ is the wave number of the photon which is small compared to the electron wave numbers. The selection rule follows directly from the character of the Bloch wave function for electrons in a periodic potential field; the matrix element of dipole moment for the optical transition vanishes unless (5.1) is fulfilled. Because of the selection rule, the long wavelength threshold of the absorption may correspond to a photon energy larger than the energy gap, if the minimum of the conduction band and the maximum of the valence band are not at the same point in

k space. However, lattice vibration perturbs the periodic potential and relaxes the selection rule. The process may be considered as optical transitions accompanied by the emission or absorption of a phonon. It is commonly called indirect transitions, while allowed transitions are called direct. Due to indirect transitions, the threshold of measurable intrinsic absorption may be nevertheless related to the energy gap even if the direct transitions should begin at shorter wavelengths.

The intrinsic absorption edge usually shifts in frequency when pressure is applied to the semiconductor. This can be interpreted as a variation of the energy gap with lattice dilatation. Temperature variation also produces a shift. The effect is in part due to lattice dilatation which can be estimated from the pressure shift by using the compressibility and the thermal expansion coefficient. Another effect is that the change of carrier-lattice interaction energy, as a carrier is lifted from the valence to the conduction band, is temperature dependent. This effect may be more important; the observed temperature shift is often much larger or even opposite in sign as compared to the lattice dilatation effect. In an optical transition, the state of lattice vibration remains unchanged, according to the Frank-Condon principle. On the other hand, thermal excitation of a hole-electron pair is associated with a change in the state of lattice vibration, as the free energy of the system is minimized for the initial as well as the final states. However, the quantity E_a determining the intrinsic carrier concentration is the free energy for hole-electron production, which turns out to be approximately the same as the energy determining the threshold for intrinsic absorption.

As an example, we may consider the intrinsic absorption edges of germanium and silicon, which have been studied in detail. Pure crystals show very little absorption at long wavelengths and intrinsic absorption near the threshold can be studied down to very low levels. The fact that the near-threshold absorption corresponds to indirect transitions is established by the fact that with decreasing temperature the curve not only shifts to shorter wavelength, but also decreases in magnitude, in agreement with expected diminution of indirect transition with decrease of lattice vibration. Furthermore, the square root of the absorption coefficient plotted against frequency shows a tail at low levels which may be interpreted as transitions involving phonon absorption rather than phonon emission; the tail diminishes rapidly with temperature as may be expected. In the case of germanium, a sharp turn of the absorption at ~ 1.5 μm indicates possibly the onset of direct transitions. The absorption threshold and its variation with pressure and temperature have been determined, the variation being of the order of magnitude expected from theoretical calculations. The optical data are in good agreement with the information about E_a obtained from electrical data on the intrinsic carrier concentration.

5·2.1. Effect of strong magnetic fields. – It has been found recently that the transmission inside the threshold of intrinsic absorption shows oscillations as the wavelength is varied, when a strong magnetic field is applied. The effect has been observed in InSb [19], Ge [20] and InAs [20]. The absorption edge itself shifts to shorter wavelengths. The effect is attributed to the quantization of electronic energy levels in a magnetic field which is the basis of Landau's theory of carrier diamagnetism. In the presence of a magnetic field along the z direction, the energy levels are given by

$$E = \frac{\hbar^2 k_z^2}{2m} + \frac{e\hbar}{mc} H\left(n + \frac{1}{2}\right) = E_z + \frac{e\hbar}{mc} H\left(n + \frac{1}{2}\right), \tag{5.2}$$

where n is an integer. The density of states, number of states per unit energy range, varies as $1/\sqrt{E_z}$ for a given n; it has, therefore, high peaks corresponding to successive values of n. Optical transitions between two energy bands, each having a spectrum of this character, may be expected to show oscillations with peaks at

$$h\nu = h\nu_t + \frac{e\hbar H}{c}\left(\frac{n_c + \frac{1}{2}}{m_c} + \frac{n_v + \frac{1}{2}}{m_v}\right), \tag{5.3}$$

where $h\nu_t$ is the absorption threshold for $H = 0$ and subscripts c and v refer to the conduction and valence bands, respectively.

The effect was observed in InSb by Burstein and Picus. Three oscillations were observed in the transmission curve under applied fields up to 20 000 G. The first minimum was attributed to $n_c = n_v = 0$. The second and third minima were assumed to correspond to ($n_c = 0$, $n_v = 1$) and ($n_c = 1$, $n_v = 0$), respectively. According to this assignment, the data yield $m_c = 0.016$ m for the conduction band and $m_v = 0.044$ m for the valence band. It should be mentioned that the valence band of InSb is believed to consist of two overlapping branches, one of which having an estimated $m = 0.12$ m plays the important role in conduction processes. This branch of the valence band did not show its effect in the oscillation, probably because of the heavy mass. Zwerdling and Lax observed the effect in germanium by using magnetic fields up to 36 000 G. Fig. 7 shows one of the curves obtained. The oscillations were observed in the region of direct transitions, between 1.4 and 1.6 μm. The frequencies of various minima

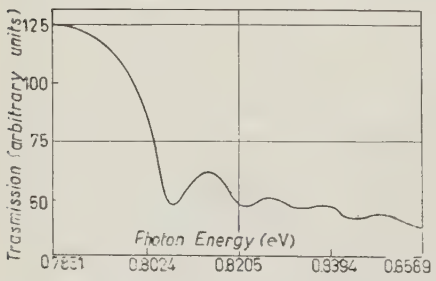


Fig. 7. – Infrared transmission of germanium as a function photon energy, under an applied magnetic field of 36 000 gauss.

in the transmission curve vary linearly with the magnetic field, and extrapolate to a common $h\nu_t = 0.803$ eV at $H = 0$, which is assumed to be the separation of the conduction and valence bands at $k = 0$.

5·2.2. Effect of high impurity concentration. One of the simple effects of impurities on the intrinsic absorption is to change the electron occupation of the energy levels by giving the sample a large carrier concentration. Thus, if the levels near the bottom of the conduction band were occupied by conduction electrons, optical transition of electrons from the valence band to these levels will be ruled out. Such an effect has been observed in n-type InSb by many workers, and in n-type InAs by SPITZER. In samples with large carrier concentrations, the edge of intrinsic absorption is shifted to increasingly shorter wavelengths. It is readily shown that the photon energy corresponding to a given absorption coefficient α is given by

$$h\nu(\alpha) = h\nu_t + \left(1 + \frac{m}{m}\right)\left[\zeta - kT \ln\left(\frac{\alpha_0 - \alpha}{\alpha}\right)\right], \tag{5.4}$$

where α_0 is the absorption coefficient at $h\nu$ in samples of low carrier concentrations. It is assumed in this expression that the absorption is produced by direct transitions, and that the maximum of the valence band is at the same point in k space as the minimum of the conduction band. However, for $m_v \gg m_c$, the ratio m_c/m_v can be dropped and the resulting expression holds approximately without the limitations. Gobeli's measurements [21] given in Fig. 8 show that the experimental points for the sample of large carrier concentrations can be fitted with curves calculated by using this expression and the experimental data for the purer sample. The value of ζ is determined by the curve fitting, and knowing the carrier concentration, the effective mass can be estimated.

The effect and interpretation are not always so simple, and departures from simple behavior provide useful information. For example, samples

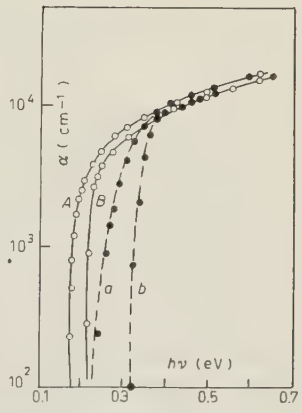


Fig. 8. – Absorption coefficient, α, of InSb as a function of photon energy. The circle points were measured on a p-type sample with a low carrier concentration; the points on A and B were measured at 298 °K and 77 °K respectively. The dot points were obtained on an n-type sample, having $6.1 \cdot 10^{17}$ cm^{-3} electron concentration at 298 °K (a), and 77 °K (b). The curves (a) and (b) were calculated according to (5.4) using $m^* = 0.03m$.

of InSb having the same carrier concentration should have the same absorption edge regardless of the type of conduction, if the conduction and valence bands are simple and the transitions are direct. However, the absorption edge of p-type samples is quite different from that of n-type samples which have the same carrier concentrations. The experimental data indicate that there is an admixture of indirect transition.

The effect of impurities may be more than merely influencing the transition probability through the addition of carriers. It may increase the probability for indirect transitions by relaxing the selection rule. Also, the energy level spectrum may be affected by the presence of impurities. These more complicated effects are indicated in some unpublished results obtained in germanium and GaSb samples by SPITZER and RAMDAS.

5·2.3. Exciton absorption. – On the longer wavelength side of intrinsic absorption edge, there may appear a series of absorption bands due to the production of excitons. The phenomena are well-known in alkali halide crystals where the bands occur in the far ultraviolet. Recently sharp and intense exciton bands have been observed in a number of other ionic crystals in or near the visible region. The investigation for Cu_2O has been very extensive and informative. The subject has been reviewed in a recent paper [22].

5·3. *Carrier effect.* – The motion of a free charge under the oscillating electric field of radiation gives a current 90° out of phase with the electric field. Energy loss is produced only as a result of collisions suffered by the carrier, which give an in-phase current component. Solving the Boltzmann equation for an ac field, we get the expression for current,

$$(5.5) \qquad j_x = \frac{2}{(2\pi)^3}\frac{e^2}{\hbar}\left[E_x\int \frac{\tau v_x}{1+(\omega\tau)^2}\frac{\partial f_0}{\partial k_x}\,\mathrm{d}\Omega_{\boldsymbol{k}} - \frac{\partial E_x}{\partial t}\int \frac{\tau v_x^2}{1+(\omega\tau)^2}\frac{\partial f_0}{\partial k_x}\,\mathrm{d}\Omega_k\right].$$

For semiconductors, usually $\omega\tau \gg 1$ in the infrared region. If τ is a constant, the conductivity, σ, which gives the absorption will be proportional to $1/\omega^2$. However, τ depends on the carrier velocity. Furthermore, in view of field quantization, the energy of infrared excited carriers is quite different from the thermal energy. Therefore, the relaxation time for low frequency conduction phenomena cannot be used directly for τ. The absorption process can be treated quantum mechanically by perturbation methods. This has been done for the two major types of scattering, lattice vibration and charged centers. The absorption varies less with frequency than $1/\omega^2$ for scattering by lattice vibration and close to $1/\omega^{3.5}$ for scattering by charged centers. Experimentally, careful studies of free carrier absorption have been made only for n-type germanium. Satisfactory agreements were obtained with the theory.

The out-of-phase current which gives the susceptibility is independent of the collision time, if $\omega\tau \gg 1$. The susceptibility can be expressed in the simple form [23]

$$\chi = \frac{Ne^2}{m_s\omega^2}\,, \tag{5.6}$$

where m_s is related to the effective mass parameters:

$$m_s = m^*\,,$$

if the effective mass is a constant scalar,

$$m_s = 3(1/m_1 + 1/m_2 + 1/m_3)^{-1} \tag{5.7}$$

for a multi-ellipsoidal energy band,

$$m_s = (m_1^{\frac{3}{2}} + m_2^{\frac{3}{2}})/(m_1^{\frac{1}{2}} + m_2^{\frac{1}{2}}) \tag{5.8}$$

for an energy band with two branches which meet at the band edge, and

$$m_s = \hbar^2 k_\zeta \Big/ \left(\frac{\partial E}{\partial k}\right)_\zeta\,, \tag{5.9}$$

for degenerate carriers the energy of which depends on the magnitude of k. Thus the susceptibility gives an estimate of the effective mass. The susceptibility can be obtained from the two optical constants, refractive index, n and extinction coefficient, k,

$$4\pi\chi \equiv \varepsilon - \varepsilon_0 = n^2 - k^2 - \varepsilon_0\,, \tag{5.10}$$

where ε_0 is the dielectric constant of the crystal without the effect of carriers. The optical constants can be determined by measuring the reflection

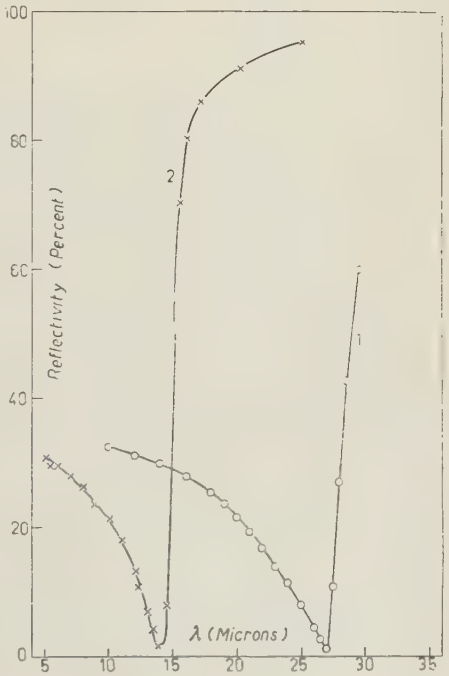


Fig. 9. – Reflectivity as a function of wavelength for two n-type samples of InSb. The carrier concentration is $6.2 \cdot 10^{17}$ cm^{-3} in case (1) and $4.0 \cdot 10^{18}$ cm^{-3} in case (2).

and transmission of radiation. Such measurements have been made on a number of semiconductors [23]. A sample with sufficient carrier concentration shows a reflectivity which with increasing wavelength goes through a minimum and rises to high values. The behavior is illustrated by the curves in Fig. 9. For germanium and silicon, the values of m_s obtained are in agreement with values expected from cyclotron resonance data. For n-type InSb, the effective mass was found to be energy dependent. Measurements have also been made on tellurium, using polarized radiation to obtain components of the mass tensor parallel and perpendicular to the optical axis.

5·3.1. Cyclotron resonance. Cyclotron resonance absorption has been observed at infrared frequencies. In order to observe the resonance absorption, it is necessary to use a frequency high as compared to the collision frequency of the carriers. For this reason, it may be necessary to use infrared rather than microwave frequencies. With higher frequency, higher magnetic field is required, according to $\omega = eH/m^*C$. The required field is reduced proportionately for carriers of small effective mass. Resonance absorption was observed in n-type InSb by Burstein, Picus and Gebbie [24] at 41.1 μm, using dc magnetic field up to 60 000 G. Using pulsed magnetic field up to 320 000 G, Keyes *et al.* [25] observed resonance absorption in Bi and n-type InSb and InAs, in the wavelength region from 10 to 22 μm. In all these cases, the carrier effective mass is small, being of the order of a few hundredth of the free electron mass.

5·3.2. Interband transitions. – Carriers may also give rise to interband transitions. Conduction electrons may be excited to a higher energy band and holes may make transitions to a lower energy band. In p-type germanium, the holes give three absorption bands associated with transitions between three overlapping branches of the valences band. An absorption band associated with interband transitions of holes has been observed in tellurium at 10 μm; the absorption is dichroic, being present only for light polarized parallel to the optical axis. Recently, an absorption band with a peak at 2.3 μm has been observed in n-type silicon [26]; the band is associated with the excitation of conduction electrons to a higher band. This type of absorption seems to be fairly common and provides useful information about the energy band structure.

5·4. *Impurity effect.* – It is well known that absorption bands associated with defect centers in insulating crystals can be produced by X-ray or electron irradiation. In semiconductors, high energy nucleons and electrons produce lattice defects by collisions with the atoms. An absorption band associated with the defects has been observed at 1.8 μm in irradiated silicon. Infrared

absorption associated with the ionization of various impurities has been observed in both germanium and silicon. In the latter material, impurity excitation bands were observed which are very helpful for the study of impurity states. This problem will not be discussed further, since it will be covered in other lectures.

Impurity effect may show up in the radiation which is produced by the recombination of excess electrons and holes. Such radiation which is intrinsic property of the crystal was first observed in germanium. HAYNES [27] observed in silicon samples, which were cooled to 77 °K, that there were additional structures on the long wavelength side of the peak of intrinsic radiation. The excess carriers were produced by passing a current pulse through a p-n junction in the specimen, and radiation from the neighborhood of the junction was observed. At the low temperature, some of the donor or acceptor impurities became unionized and their recombination with the excess carriers gave rise to the peaks which differ from the intrinsic peak by the ionization energies of the corresponding impurities. Some evidence of similar effect in germanium has been reported by AIGRAIN [28].

5·5. *Lattice vibration.* – Lattice vibration waves which produce transverse polarization in the crystal can be excited by radiation through the interaction between the electric field and the polarization. The strength of excitation depends on the effective charge of the vibrating atoms. The absorption associated with lattice vibration can therefore be used to determine the effective charge. Example of such work are the recent studies of InSb [29] and various other III-V compounds [30]. These semiconductors have zinc blende type lattice similar to the diamond type lattice of the monatomic valence crystals, germanium and silicon, and it is of special interest to determine the nature of binding, homopolar or ionic. It turns out that these compounds all show a strong absorption and reflection band in the 30-60 μm region, e.g., 5.3 μm in InSb. The effective charge e^* is given by the expression

$$\varepsilon_0 - \varepsilon_\infty = \frac{4\pi N e^*}{M\omega_0^2}\left(\frac{\varepsilon_\infty + 2}{3}\right)^2,$$

where ε_0 and ε_∞ are respectively the dielectric constants for frequencies much lower and much higher than the vibration frequency ω_0 which corresponds to peak absorption and M is the reduced mass of the atoms. ε_0 and ε_∞ can be determined from reflectivities at appropriate frequencies. The value of e^* obtained is $0.34\,e$ for InSb and varies from $0.18\,e$ (InP) to $6.08\,e$ (GaSb) for the other compounds.

The effect of impurity on lattice absorption is playing an important role in the investigation of oxygen in silicon. Both germanium and silicon have

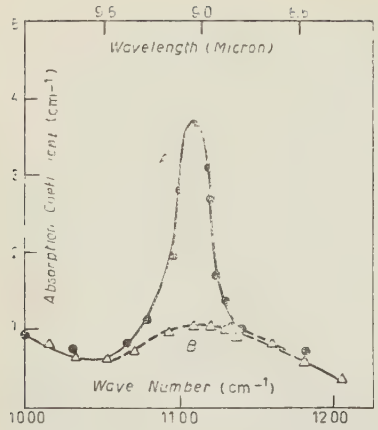


Fig. 10. – Infrared absorption of silicon. *A*) For sample pulled from a quartz crucible. *B*) For sample prepared by floating zone technique.

a number of rather weak lattice absorption bands which are temperature dependent, becoming weaker as the temperature is lowered. In silicon, the absorption band at 9 μm is an exception. KAISER [31] showed that this band is associated with the presence of oxygen and its strength can be correlated with the oxygen content determined by vacuum fusion analysis. Oxygen can come from the quartz crucible used for the melt from which the crystal was pulled, and low oxygen content was found in crystals prepared by the floating zone technique. Fig. 10 shows the absorption band for two samples prepared by the two methods. The wavelength of the absorption band is close to that of the absorption in SiO_2 films, 9.3 μm. The band was attributed to Si—O bond stretching vibration, assuming that two neighboring silicon atoms give up their covalent bond to bind with an interstitial oxygen atom. According to this interpretation, an associated band may be expected at a longer wavelength, and it has been suggested that this is the band observed at 19.4 μm [32]. In germanium, KAISER observed that as the lattice bands became generally weaker with decreasing temperature a band became more pronounced at 11.6 μm as shown in Fig. 11. This absorption was attributed to Ge—O bond stretching vibrations. It was shown that the hexagonal form of GeO_2 has a very strong absorption band at 11.5 μm.

Recently, SPITZER observed a new absorption band at 20.5 μm in silicon samples irradiated by reactor neutrons. The band is close on the long wavelength side of the lattice bands in ordinary silicon. The strength of the band increases with increasing neutron

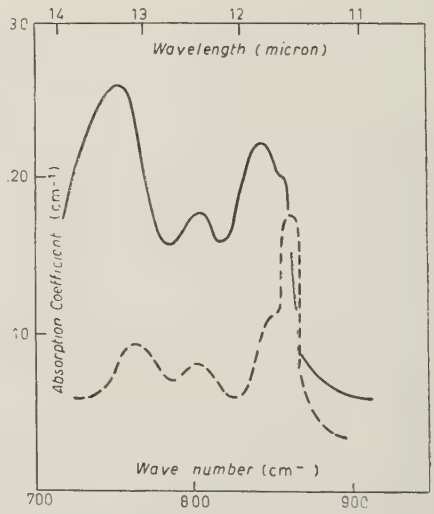


Fig. 11. – Infrared absorption in germanium sample pulled from a quartz crucible. Solid curve was measured at 300 °K and the dashed curve was measured at 90 °K.

flux, indicating that the band is due to the effect of neutron-produced defects on the lattice vibration.

5·6. *Inhomogeneity effects.* – Dislocations in crystals can be made visible by « decorating » them with aggregates of some kind of atom. The method has been used for silver bromide [33] in which the dislocation was decorated with photolytic silver, and for additively colored sodium chloride [34] in which the excess of sodium ions served the purpose. Dash [35] applied the method to silicon which gives good transmission for wavelengths longer than the intrinsic absorption threshold at 1.1 μm. An infrared image tube was used for visual observation and infrared plates were used for taking photographs. Copper was used for the decoration. Its high diffusion rate in silicon and decreasing solubility with temperature below 1200 °C facilitate precipitation upon cooling to room temperature. The specimens were etched in a suitable solution which develops pits at points where dislocations meet the surface. Following the etching, a small amount of metallic copper or $Cu(NO_3)_2$ solution was placed on the specimens, and the latter were heated in a hydrogen atmosphere at 900 °C for an hour. The specimens were viewed or photographed by transmitted light. The dislocations show up clearly as dark lines. Crystals as grown from the melt and plastically deformed specimens were studied. The dislocations were found to have a strong preference for ⟨110⟩ direction, and they connect etch pits on the surfaces. Concentric loops corresponding to the pattern associated with volume sources were also observed.

Bond [36] obtained infrared photographs of edge dislocations in silicon using the photoelastic effect. The crystal becomes birefrangent in the stress field around a dislocation. The stress field can thus be photographed by using polarized light and crossed Nicols. Photographs were obtained which show patterns corresponding to the stress field of a single dislocation and having extensions of 10 to 20 μm. The method is effective for seeing edge dislocations «end on » or at a small angle. Computations show that screw dislocations cannot be photographed along the axis and photographs of either dislocation in the perpendicular direction would require very long exposures.

Light scattering produced by aggregates of foreign atoms was observed by Kaiser [37] in studying the effect of oxygen in silicon. It has been mentioned that the presence of oxygen increases the lattice absorption band at 9 μm. It was found that heat treatment of silicon at 1000 °C reduced the 9 μm band without changing the oxygen content as determined by vacuum fusion analysis. At the same time, the intrinsic absorption edge developed a tail with a $\lambda^{-3.4}$ dependence. It was shown that the tail was produced by Rayleigh scattering; infrared light scattered by a square sample in a direction perpendicular to the direction of incident white light was detected by using an infrared image tube. The observed wavelength dependence of the tail is

reasonably close to λ^{-4} expected for the scattering. The scattering was attributed to clusters of oxygen atoms. After heat treatment at 1350 °C, the scattering was reduced and the strength of the 9 μm absorption band was restored.

REFERENCES

[1] H. H. Woodbury and W. W. Tyler: *Phys. Rev.*, **105**, 84 (1957); **1-2**, 697 (1956)
[2] C. Herring and E. Vogt: *Phys. Rev.*, **101**, 944 (1956).
[3] C. S. Smith: *Phys. Rev.*, **94**, 42 (1954); R. W. Keyes: *Phys. Rev.*, **100**, 104 (1955); F. J. Morin, T. J. Geballe abd C. Herring: *Phys. Rev.*, **105**, 525 (1957).
[4] H. Fröhlich: *Advances in Physics*, **3**, 325 (1954).
[5] H. Ehrenreich and A. W. Overhauser: *Phys. Rev.*, **104**, 331, 649 (1956).
[6] N. Sclar: *Bull. Am. Phys. Soc.*, **1**, 48 (1956); *Phys. Rev.*, **104**, 1548 (1956); F. Blatt: *Bull. Am. Phys. Soc.*, **1**, 48 (1956); *J. Phys. Chem. Solids*, **1**, 262 (1957).
[7] N. Sclar: *Phys. Rev.*, **104**, 1559 (1956).
[8] J. A. Swanson: *Phys. Rev.*, **99**, 1799 (1955).
[9] For a general discussion of the problem, see H. M. James: *Purdue Semiconductor Research Quarterly Report*, Oct. 1 to Dec. 31, (1956).
[10] The latest measurement was reported by R. M. Broudy and J. D. Venables: *Phys. Rev.*, **105**, 1757 (1957). For theory see C. Herring and E. Voigt: *Phys. Rev.*, **101**, 944 (1956).
[11] J. W. McClure: *Phys. Rev.*, **101**, 1642 (1956).
[12] J. A. Swanson: *Phys. Rev.*, **99**, 1799 (1956).
[13] For recent work see P. N. Argyres and E. N. Adams: *Phys. Rev.*, **104**, 900 (1956).
[14] H. P. R. Frederickse and W. R. Hosler: *Canad. Journ. Phys.*, **34**, 1377 (1956).
[15] R. W. Keyes and R. J. Sladek: *Phys. Rev.*, **100**, 1262A (1955).
[16] C. Goldberg, E. N. Adams and R. E. Davis: *Phys. Rev.*, **105**, 865 (1957).
[17] A comprehensive survey has been given by C. Herring: paper presented at International Conference, Garmisch 1956 (in print).
[18] A more detailed discussion with extensive references has been given by the author in a previous article, *Infrared Absorption in Semiconductors*, in *Rep. Progr. Phys.* (London, Physical Society), **19**, 107 (1956). References will be given here only to more recent works.
[19] E. Burstein and G. S. Picus: *Phys. Rev.*, **105**, 1123 (1957).
[20] S. Zwerdling and B. Lax: *Phys. Rev.*, **106**, 51 (1957); B. Lax, S. Zwerdling and L. M. Roth: *Bull. Am. Phys. Soc.*, **2**, 141 (1957).
[21] H. Y. Fan and G. W. Gobeli: *Bull. Am. Phys. Soc.*, **1**, 111 (1956); G. W. Gobeli and H. Y. Fan: *Bull. Am. Phys. Soc.*, **2**, 121 (1957).
[22] E. F. Gross: *Suppl. Nuovo Cimento*, **3**, 672 (1956).
[23] W. G. Spitzer and H. Y. Fan: *Phys. Rev.*, **106**, 882 (1957); for tellurium see R. S. Caldwell and H. Y. Fan: *Bull. Am. Phys. Soc.*, **2**, 135 (1957).
[24] E. Burstein, G. S. Picus and H. A. Gebbie: *Phys. Rev.*, **103**, 825 (1956).

[25] R. J. Keyes, S. Zwerdling, H. H. Holm and B. Lax: *Phys. Rev.*, **104**, 1804 (1956).
[26] W. G. Spitzer and H. Y. Fan: *Phys. Rev.*, **108**, 268 (1957).
[27] J. R. Haynes and W. C. Westphal: *Phys. Rev.*, **102**, 1676 (1956).
[28] P. Aigrain: *Physica*, **20**, 1010 (1954).
[29] W. Spitzer and H. Y. Fan: *Phys. Rev.*, **99**, 1893 (1955).
[30] G. S. Picus and E. Burstein: *Bull. Am. Phys. Soc.*, **2**, 66 (1957).
[31] W. Kaiser, P. H. Keck and C. F. Lange: *Phys. Rev.*, **101**, 1264 (1956).
[32] H. G. Hrostowski and R. H. Raiser: *Bull. Am. Phys. Soc.*, **1**, 295 (1956).
[33] D. N. Hedges and J. W. Mitchell: *Phil. Mag.*, **44**, 233 (1953).
[34] S. Amelinckx: *Phil. Mag.*, **1**, 269 (1956).
[35] W. C. Dash: *Journ. App. Phys.*, **27**, 1193 (1956).
[36] W. L. Bond and J. Andrus: *Phys. Rev.*, **101**, 1211 (1956).
[37] W. Kaiser: *Phys. Rev.*, **105**, 1751 (1956).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Semiconducting Compounds.

G. A. BUSCH

Swiss Federal Institute of Technology - Zürich

I.

General considerations.

1. – Historical remarks.

Research on semiconducting compounds started 50 years ago when BAEDECKER observed semiconducting properties in CuI, while investigating its equilibrium in the presence of iodine-vapour. Much work was done subsequently by GUDDEN, KOENIGSBERGER, JOFFÉ and others dealing generally with oxydes, sulphides and selenides.

For a long time the physics of semiconducting compounds had a rather bad reputation because of the importance of chemistry in the reproducibility of its results. The chemical equilibrium with oxygen for instance is most important for compounds like ZnO, Cu_2O and other oxides. It is well known that ZnO is an excess semiconductor (conduction by electrons) while Cu_2O is a deficit semiconductor (conduction by holes). A third case, also well known, is that of amphoteric semiconductors like PbS. With exactly stöchiometric composition PbS is a poor conductor, however, its electrical conductivity increases rapidly with increasing as well as with decreasing sulphur-concentration. Renewed interest in the relation between chemical reactions and semiconductor properties was shown in recent years.

A few years ago an extremely important field of research has been opened by the observation of the so-called « semiconducting intermetallic compounds ». Today it seems that the term intermetallic is not really adequate, since no semiconductor is a compound of two true metals, unless one of them has some sort of allotropic non-metallic modifications.

2. – Semiconductivity and chemical bond.

The distinction between insulators and semiconductors is quite arbitrary. A more important distinction has to be made, however, between solid ionic conductors and semiconducting compounds. For a compound to be called a semiconductor the electronic conductivity σ_e should be much larger than the ionic conductivity σ_i. Since in general the mobilities $\mu_e \gg \mu_i$, the energy to release an ion must be larger or at least equal to that necessary to release an electron: $Q_i \geqslant \Delta E$. As a rule Q_i is of the order of 1 to 2 eV. One would expect therefore substances with intrinsic activation energies ΔE higher than about 2 eV to be ionic conductors. For lower values of ΔE electronic conductivity should be predominant. This change of the character of conductivity is shown in Fig. 1 for the isoelectronic sequence Sn, InSb, CdTe and AgI. Room temperature values for the intrinsic conductivity are plotted against the difference of atomic number of the components, which means increasing ionicity. CdTe is still a typical semiconductor whereas in the case of AgI ionic conductivity predominates. Experimentally the distinction between electronic and ionic conductivity and the determination of the ratio of the two contributions is often difficult. Careful measurements of transport numbers are necessary.

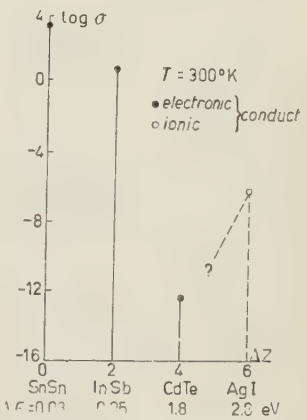


Fig. 1. Intrinsic electronic conductivity compared with ionic conductivity for iso electronic compounds.

A still more important problem is the question whether a given compound is metallic or non-metallic, i.e. semiconductive. At the present time there is no theory which allows in a strict manner to predict metallic or non-metallic conduction for a compound, even if its crystallographic structure is known. This is essentially a question of the chemical bond, which can not be answered *a priori*. If one allows Mn to react with Te, for example, it is impossible to predict what kind of compound will be formed. All one can do so far is to grow the crystal and to try to explain its properties *a posteriori*. There are some interesting rules, however, which shall be discussed here. A first rule is due to ZINTL: For a binary compound to be non-metallic at least one component must have a valence equal to or larger than 4. Table I shows a list of compounds of Mg where the first three columns represent metallic compounds, and the last four represent semiconductors with the structure indicated at the bottom of the respective columns. The structures indicated are not exclusively metallic

or non-metallic. There are cases like that of $AuIn_2$, which has the structure of CaF_2 but is a metal.

Table I.

I	II	III	IV	V	VI	VII
		Mg_4Al_3	Mg_2Si	Mg_3P_2	MgS	$MgCl_2$
		MgAl				
		Mg_3Al_4				
		$MgAl_3$				
Mg_2Cu	Mg_7Zn_3	Mg_5Ga_2	Mg_2Ge	Mg_3As_2	MgSe	$MgBr_2$
$MgCu_2$	MgZn	Mg_2Ga				
	Mg_2Zn_3	MgGa				
	$MgZn_2$	$MgGa_{1+x}$				
	Mg_2Zn					
Mg_3Ag	Mg_3Cd	Mg_5In_2	Mg_2Sn	Mg_3Sb_2	MgTe	MgJ_2
MgAg	MgCd	Mg_2In				
	$MgCd_3$	MgIn				
		$MgIn_2$				
Mg_3Au	Mg_3Hg	Mg_5Tl_2	Mg_2Pb	Mg_3Bi_2		
Mg_5Au_2	Mg_2Hg	Mg_2Tl				
Mg_2Au	MgHg	MgTl				
MgAu	$MgHg_2$					
Characteristic structures for metals and alloys			CaF_2-type	Mn_2O_3- and La_2O_3-type	NaCl- and ZnS-type	$CdCl_2$- and CdJ_2-type
			Characteristic structures for salt-like compounds.			

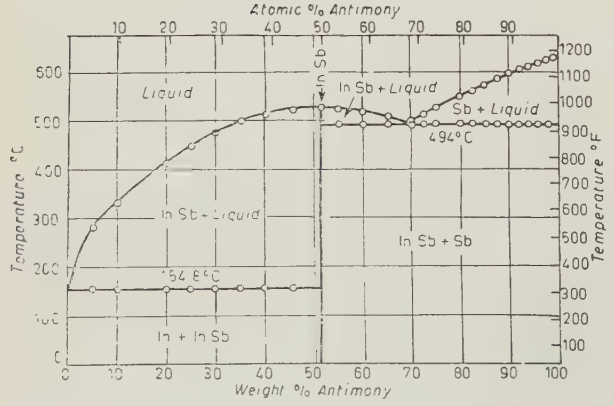


Fig. 2. – Phase diagram of the alloy-system In—Sb.

The line separating column III and IV is the so-called Zintl-boundary. The compounds on the left side are really intermetallic whereas those on the right side are typical semiconductors.

Compounds show up clearly in the phase diagram. In Fig. 2 a typical example is shown. The compound InSb shows a

relative maximum of the liquidus-curve at 540 °C. It crystallizes with the zincblende structure, which is entirely different from the structures of the components In and Sb.

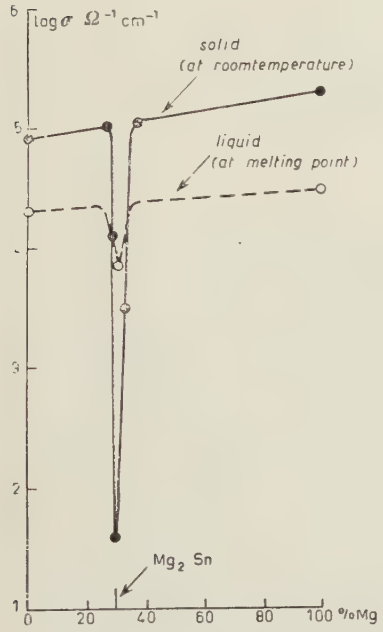


Fig. 3. – Electrical conductivity in the alloy-system Mg—Sn in the solid and the liquid state.

Measurements of the conductivity as a function of the concentration give precise information about the existence of compounds. Fig. 3 shows a very sharp notch in the curve of conductance for Mg_2Sn against concentration of Mg. This minimum is evident even in the liquid state (dotted curve) and is probably due to the presence of molecules in the melt.

The transition between the solid and the liquid state is connected with rather striking effects. Fig. 4 shows the variation of the conductivity and the Hall-coefficient for InSb in the solid and the liquid state as a function of temperature.

Below the melting point we find positive values for $d\sigma/dT$, above the melting point $d\sigma/dT$ is negative, however. The Hall coefficient is large and decreases with temperature in the solid state, while it is small and practically constant in the liquid state. This means that a complete breakdown of the lattice occurs in the liquid accompanied by an increase of co-ordination number. The liquid state is practically metallic with a very low Hall effect due to the high concentration of electrons. In contrast Mg_2Sn shows increasing conductivity also in the liquid state, which can probably be explained by progressive dissociation of the molecules in the melt, the latter becoming more and more metallic.

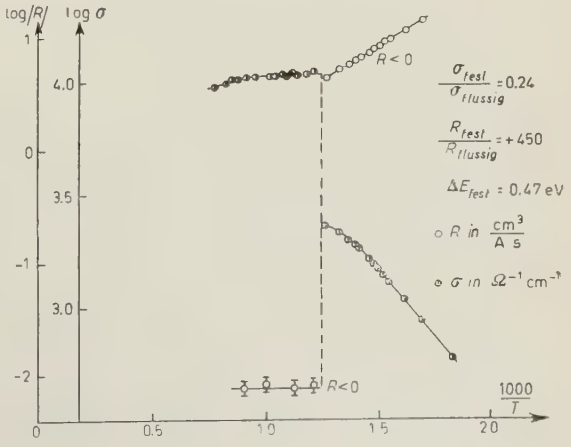


Fig. 4. – Electrical conductivity and Hall-coefficient for InSb in the liquid and the solid state.

The question, whether a compoound is a metal or a semiconductor, has been studied very recently by MOOSER and PEARSON [1] by extending Pauling's valence-bond theory of metals to the more appropriate case of semiconductors. These authors emphasize the importance of the covalent bond for the formation of non-metallic structures. If a compound is to be a semiconductor, at least one atom must have completely filled *s*- and *p*-shells. For elementary semiconductors like Si and Ge this condition is the same as the well known octet-rule of moleuclar chemistry. It means for example that each Ge-atom can form covalent bonds with its four nearest neighbours in the lattice.

The situation in a compound is more complicated. In the case of InSb for instance, there are two possible types of chemical bonds. In one of these one electron shifts from the Sb atom to the In atom, thus creating In^- and Sb^+ atoms, each with four valence electrons. The formation of covalent bonds necessary to explain the zincblende structure of InSb is thus possible. The second type has all the electrons shifted from In to Sb producing In^{3+} and Sb^{3-} ions. The bond would be completely ionic in this case. The actual chemical bond, however must be interpreted as a mixture of these two extreme types, both satisfying the condition of completely filled *s*- and *p*-shells.

As an atom cannot form more electron-pair-bonds than the number of its valence electrons, a semiconducting compound must contain at least one atom, the valency of which is equal or larger than four. This is the reason for Zintl's rule. The condition just mentioned leads to the following simple rules. If one has a compound $A_a^{(\alpha)}B_b^{(\beta)}C_c^{(\gamma)}\ldots$ of this kind and if $\alpha, \beta, \gamma, \ldots$ are the numbers of valence electrons of A, B and C ... the compound should be a semiconductor if

$$A\alpha + B\beta + C\gamma + \ldots = 8n\,,$$

where n is an integer and equal to the number of atoms with valence 4 or higher in the formula. Examples are

$In^{(3)}Sb^{(5)}$ $\quad n = 1\,,$ $\qquad$ $In_2^{(3)}Te_3^{(6)}$ $\quad n = 3\,,$

$Mg_3^{(2)}Sb_2^{(5)}$ $\quad n = 2\,,$ $\qquad$ $Hg^{(2)}In_2^{(3)}Te_4^{(6)}$ $\quad n = 4\,.$

This rule does not always hold, as, for example, in the case of CdSb. MOOSER and PEARSON have given a more general rule, namely

$$\frac{n}{n_a} + b = 8\,.$$

Here n_e is the number of valence electrons available in the chemical formula, n_a is the number of atoms whose valence is equal or greater than 4, and b is

the number of bonds formed between atoms of this kind. An examination of the lattice of CdSb indicates in this case $n_e = 7$, $n_a = 1$ and $b = 1$, which satisfies the above condition. The are no equivalent rules so far for compounds containing transition elements.

Following Mooser and Pearson's rule it is possible to derive semiconductors with more complicated structures, starting out for instance from the zincblende-structure. By doubling the elementary cell of ZnS one has

$$2\,Zn^{(2)}S^{(6)} \rightarrow Zn_2^{(2)}S_2^{(6)},$$

which still satisfies the rule. We now can replace two divalent atoms by two different atoms having, namely, valence 1 and 3. So we get the formula

$$A^{(1)}B^{(3)}X_2^{(6)}$$

like for instance

$$Cu^{(1)}Fe^{(3)}S_2^{(6)},$$

which still gives the right amount of 16 electrons and is proved to exist and to be a semiconductor. Such ternary compounds with chalcopyrite structure were first investigated by Goodman and Douglas [2]. If one again doubles the elementary cell one obtains $Cu_2Fe_2S_4$, where we can replace Cu_2 by one divalent atom, thus obtaining a new compound of the type $A^{(2)}B_2^{(3)}X_4^{(6)}$.

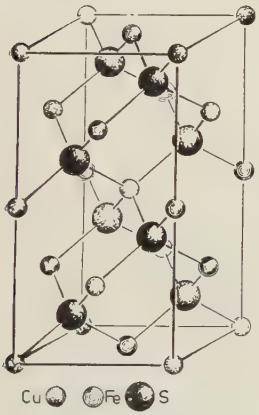


Fig. 5. – The chalcopyrite structure.

An example is

$$Hg^{(2)}In_2^{(3)}Te_4^{(6)}$$

which is again a really existing semiconductor [3] with the structure of the chalcogenides [4].

In Fig. 5 the chalcopyrite structure is shown. It can be readily derived from the zincblende-structure by putting two elementary cells on top of each other, which leads to a tetragonal cell containing 4 molecules. Each sulphur atom is still surrounded tetrahedrally by 4 atoms, but only two of them being equal. As Fig. 6 shows the chalcogenide structure is closely related to the foregoing and we arrive at it

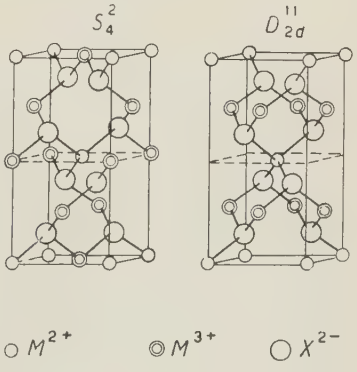


Fig. 6. The chalcogenide structure.

in a similar way. The elementary cell contains only two molecules and it can be clearly seen that two atoms are missing in the basis of the structure.

With the following components

Zn Al S

Cd Ga Se

Hg In Te

one gets 23 possible compounds of the same structure, and all of them should be semiconductors. This has been confirmed for the following compounds with intrinsic activation energies ΔE_i as indicated,

$ZnIn_2Se_4$	2.6 eV
$HgIn_2Se_4$	0.6 »
$ZnIn_2Te_4$	1.4 »
$CdIn_2Te_4$	0.9 »

The preparation of all of these compounds needs some careful techniques. The common practice of melting together the components can be applied only in a few cases. Completely sealed-off systems to prevent volatile components from escaping are often necessary. Growth by reaction from the vapour phase also may be possible. The standard technique of crystal growth and zone melting may be used. Controlled atomsphere of the volatile components is necessary in most cases.

3. – Characteristic data of some compounds.

Table II shows the characteristic data so far known of compounds of the type $A^{II}B^{VI}$ with zincblende-structure. It appears that the activation energy ΔE_i decreases systematically when going from the lighter to the heavier components.

It has to be pointed out that Hg-compounds, like HgSe and HgTe show an extraordinarly high mobility b_n of the electrons, which is one of the most striking features of these compounds. Furthermore CdS and CdSe are interesting photoconductors.

TABLE II.

	Zn	Cd	Hg
S	$\Delta E_i =$ 3.7	2	—
	$b_n =$ —	210	—
	$b_p =$ —	—	—
	$T_s =$ 1850	1750	583
Se	$\Delta E_i =$ 2.6	1.8	1
	$b_n =$ 100	—	10000
	$b_p =$ —	—	—
	$T_s =$ 1000	1350	690
Te	$\Delta E_i =$ 2.2	1.45	1
	$b_n =$ —	300	10000
	$b_p =$ —	100	100
	$T_s =$ 1239	1045	670

ΔE_i: intrinsic activation energy (eV)
T_s: melting point (°C)
b_n: mobility of the electrons (cm²/Vs)
b_p: mobilitity of the holes.

Table III shows some data about compounds with CaF_2 structure. The activation energies again decrease with the atomic weight of the components. For Mg_2Pb ΔE_i is unmeasurably small and it therefore seems to be a metal. This is indeed surprising, because this compound has exactly the same crystal structure as the others in the table, so that it is perhaps preferable to consider

TABLE III.

	Melting point (°C)	Lattice-constant (Å)	ΔE_0 (eV)	a (eV/°K)	$f_n f_p$	f_n	f_p
$MgSi$	1102	6.338	0.77	$-6.4\cdot10^{-4}$	3.8	2.75	1.38
Mg_2Ge	1115	6.378	0.74	$-7.6\cdot10^{-4}$	8.3	3.96	2.1
Mg_2Sn	775	6.760	0.36	$-2.8\cdot10^{-4}$	0.17	0.44	0.38
Mg_2Pb	550	6.794	0				
Ca_2Si			1.9				
Ca_2Sn			0.9				
Ca_2Pb			0.46				

it as a highly degenerate semiconductor. Since we know nothing about the charge distribution in the crystal, it is hard to explain this anomaly.

A most interesting case is represented by Ag_2Se and Cu_2Se, which at low temperature actually are semiconductors. In Fig. 7 the position of the Se-atoms in the lattice of Ag_2Se is shown. It is tetragonal below 130 °C and cubic above. During this transition the crystal expands very little in one direction, but the lattice parameters are only slightly changed.

Fig. 8 shows the conductivity plotted against temperature and at the above mentioned temperature one notices a sharp discontinuity in the curve corresponding to a

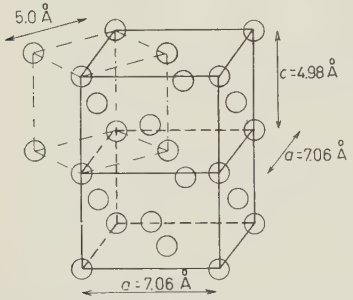


Fig. 7. – The positions of the Se-atoms in the elementary cell of Ag_2Se below and above the transition temperature.

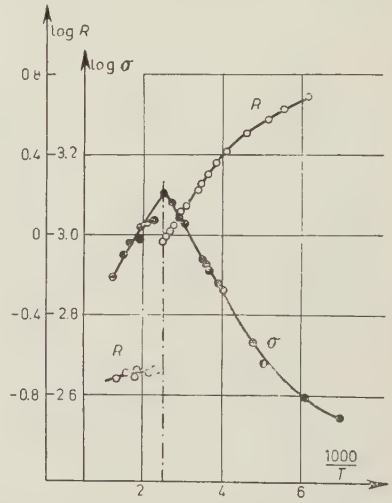


Fig. 8. – Electrical conductivity and Hall-coefficient of Ag_2Se as a function of temperature.

transition from a semiconductig to a metallic state. The same happens to the Hall coefficient.

In order to get these results, the samples had to be prepared very carefully, and the temperature had to be changed very slowly. If these conditions are not observed, results may be completely unreliable, and it seems that this is why this effect was discovered only recently [5].

BIBLIOGRAPHY

L. Pincherle and J. M. Radcliffe: *Phil. Mag.*, Suppl. **5**, No. 19, 271 (1956).
H. Welker: *Erg. Exakt. Naturwiss.*, **29**, 275 (1956).
F. A. Kröger und H. J. Vink: *Solid State Physics*, **3**, 310 (1956).

II.

Special properties of semiconducting compounds.

1. – Magnetic susceptibility.

As a first approximation the magnetic susceptibility of a semiconductor can be written as a sum of the following terms:

$$\chi = \chi_{\text{core}} + \chi_{\text{val.}} + \chi_f + \Delta\chi \,.$$

The first term is the contribution due to the electrons of the core of the atoms involved in the crystal, and is given by the Langevin formula:

$$\chi_{\text{core}} = -\frac{Ne^2}{6mc^2} \sum_{k=1}^{Z} \overline{r_k^2} < 0 \,,$$

N being the number of atoms per mole and Z the number of electrons in the core. $\overline{r_k^2}$ is the mean square distance of an electron from the centre of the core.

The second term is the contribution due to the valence electrons, and cannot be well determined, because one ignores the exact charge distribution of the valence electrons. Roughly one can write a similar formula:

$$\chi_{\text{val.}} = -\frac{Ne^2}{6mc^2} \sum_{i=1}^{V} \overline{r_i^2} < 0 \,,$$

r_i being the average radius of the orbits, and V the number of valence electrons. Both contributions are negative.

χ_f due to free charge carriers depends on the temperature. For the intrinsic range it can be shown that

$$\chi_f = aT^{\frac{1}{2}} \exp\left[-\frac{\Delta E_i}{2kT}\right] .$$

The term $\Delta\chi$ due to impurities and lattice imperfections depends in a rather complicated manner on the nature and concentration of the lattice imperfections (donors, acceptors) and is also temperature dependent. It will not be discussed here.

Fig. 9 gives results of measurements of χ on intrinsic semiconductors. The negative value of χ is plotted against

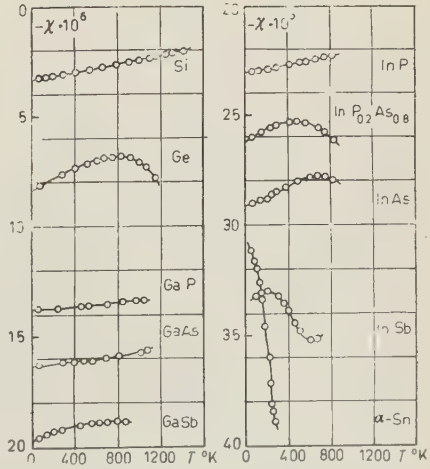


Fig. 9. – Magnetic susceptibility of several diamond-like semiconductors as a function of temperature.

temperature for different substances. The shape of the curves varies for different compounds but it can be interpreted as the sum of a linearly increasing contribution at low temperature, and a non-linear one at high temperature, the latter being due to free charge carriers. The temperature variation of χ is most pronounced for Ge, InSb and grey tin.

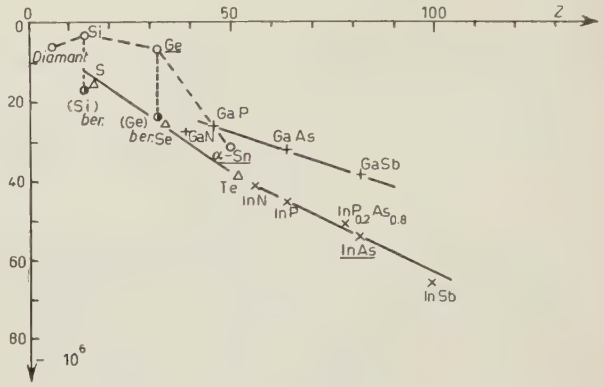


Fig. 10. – The sum $\chi_{\text{core}} + \chi_{\text{val.}}$ for several semiconductors as a function of the total number Z of electrons per molecule.

The sum $\chi_{\text{core}} + \chi_{\text{val.}}$ can be determined for pure samples at sufficiently low temperatures directly from the curves, neglecting the contribution due to the free carriers, since this diminishes very rapidly with temperature. Fig. 10 shows a plot of $-\chi$ against the number of valence electrons Z.

In Fig. 11 one sees that for isoelectric sequences $-\chi$ increases with increasing ionicity of the compounds. In each sequence the sum of the covalent radii (after PAULING) is essentially constant, but their difference varies as can be seen from the following table

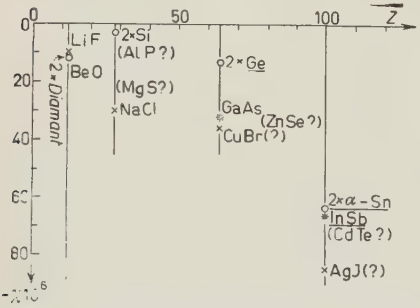


Fig. 11. – The sum $\chi_{\text{core}} + \chi_{\text{val.}}$ for isoelectronic compounds.

	$r_a + r_b$	$r_a - r_b$
SnSn	2.80 Å	0 Å
InSb	2.78 »	0.08 »
CdTe	2.80 »	0.16 »
AgI	2.80 »	0.25 »

The contribution of the atoms $\chi_a = \chi_{\text{core}} + \chi_{\text{val.}}$ can be written approximately as

$$\chi_a \cong C \cdot (\bar{r}_a^2 + \bar{r}_b^2) = \frac{C}{2} \cdot \{(\bar{r}_a + \bar{r}_b)^2 + (\bar{r}_a - \bar{r}_b)^2\},$$

$\bar{r}_a$ and $\bar{r}_b$ being the effective atomic radii. Since $r_a + r_b = \text{const}$, it follows that χ_a is proportional to the difference of the covalent radii. The results repre-

sented in Fig. 11 clearly show that in fact the experimental values for χ_a vary in the direction expected.

For a non-degenerate semiconductor in the intrinsic range the susceptibility due to the free carriers is given by the formula

$$\chi_i = n_i \frac{\beta^2}{kT}\left\{\left(1 - \frac{F_n^2}{3}\right) + \left(1 - \frac{F_p^2}{3}\right)\right\},$$

where

$$\beta = \frac{e\hbar}{2m_0 c},$$

m_0 is the rest mass of the electron, and

$$n_i = \frac{1}{4}\left(\frac{2mkT}{\pi\hbar}\right)^{\frac{3}{2}} \cdot (f_n f_p)^{\frac{3}{4}} \exp\left[-\frac{\Delta E_i}{2kT}\right],$$

represents the number of free carriers.

The positive terms represent the paramagnetic contributions due to spin-orientation both of electrons and holes, whereas the factors containing the quantities $\overline{F_p^2}$ and $\overline{F_n^2}$ represent the diamagnetic contribution. $\overline{F^2}$ is defined as follows:

$$\overline{F^2} = \overline{f_{xx} f_{yy} - f_{xy}^2}$$

where f_{rs} are the components of the effective mass tensor

$$f_{rs} = \frac{m_0}{\hbar^2} \cdot \frac{\partial^2 E}{\partial k_r \, \partial k_s}.$$

To interpret the actual value of the charge contributions one has to know the average values of the effective masses, which can be determined if the shape of the energy-surfaces is known, or by means of effective masses obtained from cyclotron resonance experiments. In the case of Ge, the parameter being well known, a rigorous computation yields results in good agreement with experiments [6]. No computations have been made for other elements and compounds so far.

Fig. 12 shows a reduced diagram of χ_f against temperature for different semiconductors. In the expression for χ_f we introduce the following variables

$$x = \frac{T}{T_0},$$

where $T_0 = \Delta E_i/2k$, and

$$y = \frac{\chi_f}{\chi_0},$$

where χ_0 contains all the constants characterizing the semiconductor. Then we get $y = -x^{\frac{3}{2}} \exp[-1/x]$.

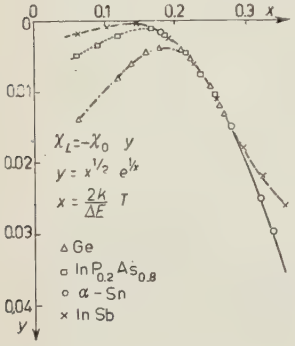


Fig. 12. – Reduced representation of the charge-carrier susceptibility for several semiconductors.

The curves in Fig. 12 show the experimental value for Ge, α-Sn and the mixed compound $InP_{0.2}As_{0.8}$. It can be seen that within a limited temperature region the results are well represented by the reduced function. There are deviations, however, at high and low temperatures. The deviation occuring in the case of InSb is partly due to degeneracy of the electron gas which is known to be strong. At low temperature the situation seems to be more complicated. A temperature dependent Van Vleck-paramagnetism as suggested by KRUMHANSL and BROOKS [2] may be responsible. On the other hand the influence of lattice defects have been definitely excluded by experiments.

2. – Thermal conductivity.

The thermal conductivity of semiconductors at high and low temperature is very important for the understanding of energy transport phenomena in solids. Theoretical discussions are given in the literature mentioned below.

Since one may assume that a semiconductor consists of an atomic lattice and an electron and hole gas, one may consider as a first approximation, that the total heat conductivity is the sum of two terms:

$$\lambda = \lambda_{\text{phonons}} + \lambda_{\text{free carriers}}.$$

This assumption is reasonable as long as scattering of phonons by electrons is not too strong.

For a non-degenerate semiconductor with a single type of carriers one may write

$$\lambda_{fc} = L \cdot \sigma_{fc},$$

where L is the Lorentz number and is given by

$$L = 2\left(\frac{k}{e}\right)^2 \cdot T.$$

As

$$\sigma_{ic} \sim \exp\left[-\frac{\Delta E_i}{2kT}\right],$$

one would expect a fast increase of λ_{fc} at high temperature. Since λ_{ph} decreases with temperature approximately as $1/T$, the total λ curve should show a minimum.

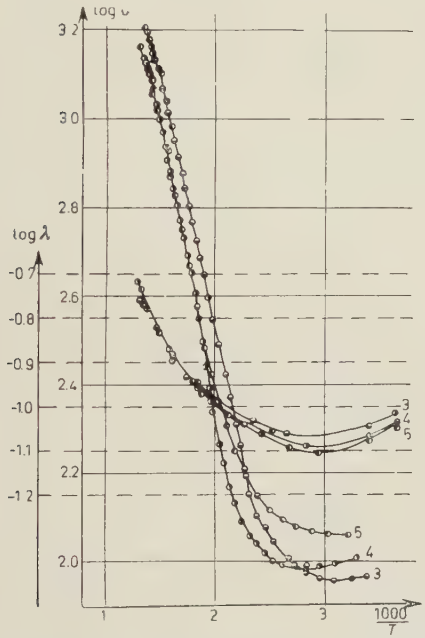


Fig. 13. – Thermal and electrical conductivity as a function of temperature for different specimens of InSb (after Busch and Schneider).

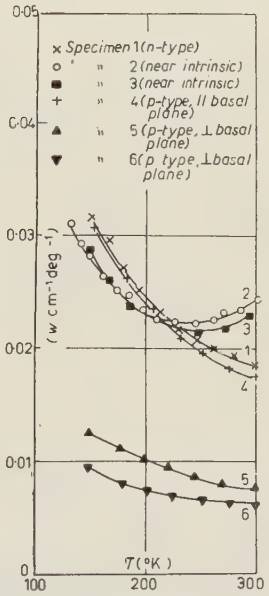


Fig. 14. – Thermal conductivity of Bi_2Te_3 as a function of temperature (after Goldsmid).

Fig. 13 shows thermal and electrical conductivity curves for different specimens of intrinsic InSb against $1/T$. One sees in fact that the conductivity increases with increasing temperature and reaches very high values.

At low temperatures λ is found to vary linearly with $1/T$. At high temperatures, however, there is a strong deviation from linearity. As can be seen from Fig. 14 for different specimens, the same effect is observed for Bi_2Te_3 which was studied by Goldsmid [8]. Curves 2 and 3 refer to nearly intrinsic specimens while 1 and 4 refer to higly doped n- and p-type specimens for conduction parallel to the basic plane of the hexagonal structure. The conductivity is smaller (curve 5 and 6) for conduction perpendicular to the basal plane.

Similar observations for PbTe, HgSe and Sb_2Te_3 have been reported by JOFFÉ. In contrast Mg_2Sn does not show this effect and there is no increase of λ with temperature (Fig. 15).

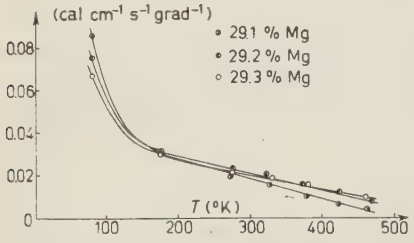


Fig. 15. – Thermal conductivity of Mg_2Sn as a function of temperature (after BUSCH and RINDERER).

An explanation may be given for the intrinsic range using the two-band model. At the high temperature end of the specimen there is a high concentration of carriers and therefore recombination of carriers occurs towards the low temperature side, which leads to an energy transport (heat flow) by pairs. This effect is typical for semiconductors and is not present in metals. Theoretical results can be obtained by proper solution of the Boltzman equation giving two equations for the electric current density and for the heat current density respectively:

$$j_x = \frac{4e}{3m}\left(AL_0 - \frac{1}{T}\frac{\mathrm{d}T}{\mathrm{d}x}L_1\right),$$

$$w_x = \frac{4}{3m}\left(AL_1 - \frac{1}{T}\frac{\mathrm{d}T}{\mathrm{d}x}L_2\right),$$

$$L_n = -\int \tau(E)\, E^n E_{\mathrm{trans}}\, D(E)\, \frac{\partial f}{\partial E}\, \mathrm{d}E\,.$$

The signs used have the following meaning:

τ:	relaxation time,
E:	total energy of electrons and holes,
E_{trans}:	translation energy of the carriers,
D:	density of states,
f:	Fermi function.

The above integral splits into two integrals limited by the band edges respectively.

The integration can be carried out by assuming a special mechanism of scattering of electrons and holes and knowing the variation of τ with energy. Setting $j_x = 0$ (i.e. no electric current flowing through the specimen) one obtains:

$$\lambda_f = \frac{4}{4mT} \cdot \frac{L_1^2 - L_0 L_2}{L_0}\,.$$

Introducing the appropriate function $\tau(E)$ for different mechanisms of scattering, one obtains for pure thermal scattering by acoustical waves for the total heat conductivity by free carriers

$$\lambda_{fc} = L(\sigma_n + \sigma_p) + 2L\,\frac{\sigma_n\sigma_p}{\sigma_n + \sigma_p}\left(\frac{\Delta E_i}{2kT} + 2\right)^2,$$

L being the Lorenz number,
σ_n the conductivity due to electrons,
σ_p the conductivity due to holes,
ΔE_i the activation energy.

The first term is due to electrons and holes, the second due to pairs. The second term is a quite substantial addition and it is interesting to see whether it explains the observed anomalies.

In the case of InSb, $\Delta E_i = 0.25$ eV at $T = 300\,^\circ$K. One has therefore $((\Delta E_i/2kT) + 2)^2 \sim 50$.

However, the factor $\sigma_n\sigma_p/(\sigma_n+\sigma_p)$ $(= \sigma_{\text{pairs}}$, because pairs do not move independently) approaches σ_p when $\sigma_n \gg \sigma_p$. Since, in fact $\sigma_n \approx 80\sigma_p$, because of the very different effective masses, the low conductivity of the holes cancels approximately the correction $((\Delta E_i/2kT)+2)^2$.

For this case one obtains approximately

$$\lambda_f \cong L(\sigma_n + \sigma_i)\left\{1 + \frac{1}{40}\left(\frac{\Delta E_i}{2kT} + 2\right)^2\right\}.$$

Thus the observed discrepancy between theory and experiment cannot be explained by the influence of recombination of pairs. For InSb the first experimental values appeared to be nearly 100 times larger than was expected. According to more recent experiments the discrepancy is much smaller but there seems to be still a factor of the order of 10 to be explained.

Apparently one has to look for a mechanism that is able to transfer heat through a semiconductor without transferring charge. That such an effect should exist has been shown by a preliminary experiment. The heat conductivity of InSb has been studied in a strong magnetic field, thus substantially reducing the carrier conductivity. Consequently one would expect a large decrease of the contribution of holes and electrons, the magneto-resistance being very large in this kind of semiconductor. In fact one observes no change of the heat conductivity within the limits of experimental error.

The following additional effects have to be considered: Firstly heat can be transfered by radiation, the cold part of the specimen receiving energy from the hot part, because the specimens are transparent in the infrared

region of the spectra. This contribution has been calculated, but it has been found to be too small to explain the effect.

A second reason is based upon the flow of excitons. As for the phonons, also an exciton-drag-effect is possible. Unfortunately too little is known about these effects at the present time, and the experiments are not yet reliable enough. Very careful experiments will be necessary with specimens containing different amounts of impurities of n and p-types.

BIBLIOGRAPHY

P. J. Price: *Phys. Rev.*, **95**, 596 (1954).
A. F. Joffé: *Can. Journ. Phys.*, **34**, 1342 (1956).

REFERENCES

[1] E. Mooser and W. B. Pearson: *Journ. Electronics*, **1**, 629 (1956).
[2] C. H. I. Goodman and R. W. Douglas: *Physica*, **20**, 1107 (1954).
[3] G. Busch, E. Mooser and W. B. Pearson: *Helv. Phys. Acta*, **29**, 192 (1956).
[4] H. Hahn, G. Frank, W. Klingler, A. D. Störger and G. Störger: *Zeits. anorg. all. Chem.*, **279**, 241 (1955).
[5] G. Busch and P. Junod (in press).
[6] Ch. Enz: *Helv. Phys. Acta*, **28**, 158 (1955).
[7] J. A. Krumhansl and H. Brooks: *Bull. Am. Phys. Soc.*, **2**, 1/3 F5 (1956).
[8] H. J. Goldsmid: *Proc. Roy. Soc.*, B **69**, 203 (1956).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2. 1958
1° Trimestre

Shallow Impurity States in Semiconductors.

W. KOHN

Carnegie Institute of Technology - Pittsburgh, Pennsylvania

In recent years there has been a rapidly increasing interest in the nature of impurity states in semiconductors. These states have a profond influence on electrical properties and also are fascinating objects for study in their own right. There are two kinds: Those whose energy levels lie somewhere near the middle of the forbidden gap, the so-called deep impurity states, and those whose levels are either very close to the conduction band or to the valence band, the shallow impurity states. At present the latter are much more completely understood and in my lectures I shall confine my attention to them. I shall touch on only a few of the many interesting studies which have been made of these states and would like to refer to a forthcoming review article for further details and references [1].

1. – Qualitative considerations.

When an impurity atom is introduced into a semiconductor, it is sometimes found that an electron can be released from this atom into the conduction band with a rather small expenditure of energy. Such an impurity is called a shallow *donor* and the weakly bound state of the electron is represented by a level just below the conduction band (See Fig. 1).

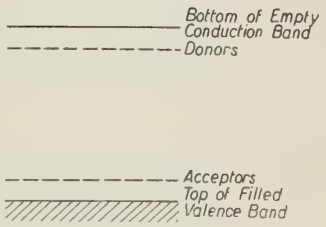


Fig. 1. – Donor and acceptor levels in the band gap of a semiconductor.

To form a qualitative picture of this state let us adopt the following imaginary procedure. We first introduce the donor minus one electron into the semiconductor. This positive ion will polarize the medium in its vicinity and, at large distances, set up an

electrostatic potential,

$$U = \frac{e}{\varkappa r}, \tag{1.1}$$

where $\varkappa$ is the static dielectric constant of the medium.

When we now introduce the missing electron it is possible that because of the exclusion principle it cannot occupy a localized orbit in the immediate vicinity of the ion, but will move in a larger orbit in which the equation (1.1) is approximately valid. As the valence band of the semiconductor is completely filled, the wave function of the impurity electron must be composed of conduction band functions and so one would expect that it satisfies a Schrödinger equation of the form

$$\left(-\frac{\hbar^2}{2m^*}\nabla^2 - \frac{e^2}{\varkappa r}\right) F = EF \tag{1.2}$$

where m^* is an appropriate effective mass of the conduction band. The solutions of (1.2) are of course the familiar hydrogenic wave functions with an effective Bohr radius of

$$a^* = \frac{\hbar^2}{m^*(e^2/\varkappa)} = a_0 \frac{\varkappa}{m^*/m_0}, \tag{1.3}$$

and an ionization energy of

$$E_0 = \frac{(e^2/\varkappa)^2}{2\hbar^2 m} = \frac{1}{\varkappa^2(m^*/m_0)} \text{ Ry}. \tag{1.4}$$

Here a_0 is the usual Bohr radius and Ry the Rydberg, 13.5 eV. In Si and Ge, because of their large dielectric constants and small effective masses, a^* is of the order of 30 Å and E_0 of the order of 0.01 eV. The lowest solution of (1.2) is

$$F = \frac{1}{(\pi a^*)^{\frac{3}{2}}} \exp\left[-r/a^*\right]. \tag{1.5}$$

It must however by no means be supposed that the function F is the complete impurity electron function. Rather, a more detailed analysis shows that when the conduction band has a single minimum at $k=0$ the impurity state wave function is a wave-packet made up of Bloch waves near $k=0$. That is it has the form

$$\psi = \sum_{\boldsymbol{k}} A(\boldsymbol{k})(\exp\left[i\boldsymbol{k}\cdot\boldsymbol{r}\right]u_{\boldsymbol{k}}(\boldsymbol{r})), \tag{1.6}$$

where $u_{\boldsymbol{k}}$ is the periodic part of the Bloch wave and $A(\boldsymbol{k})$ is zero except in the immediate vicinity of $k=0$. Because of this latter fact we can approximate $u_{\boldsymbol{k}}(\boldsymbol{r})$ by $u_0(\boldsymbol{r})$ which is identical with the Bloch wave $\varphi_0(\boldsymbol{r})$ at $k=0$. This gives

$$\psi = (\sum_{\boldsymbol{k}} A(\boldsymbol{k}) \exp [i\boldsymbol{k}\cdot\boldsymbol{r}])\varphi_0(\boldsymbol{r}) . \tag{1.7}$$

The first factor can be shown to be the $F(\boldsymbol{r})$ of Eq. (1.2), so that the total wave function becomes approximately

$$\psi(\boldsymbol{r}) = F(\boldsymbol{r})\cdot\varphi_0(\boldsymbol{r}) . \tag{1.8}$$

The two factors in Eq. (1.8) have a simple physical meaning. $\varphi_0(\boldsymbol{r})$ describes the fact that the main forces acting on the impurity electron are those of the periodic lattice in which it is embedded, so that *locally* its wave function is just that of a conduction electron at $k=0$. This has the usual nodes and rapidly varying amplitude within each atomic cell. On the other hand, the factor $F(\boldsymbol{r})$ which varies much more slowly, namely over a distance of the order a^*, describes the gentle influence of the reduced Coulomb potential, $-e^2/\varkappa r$, which holds the electron in a large orbit around the donor ion.

So far we have discussed only the donors, which can give up an electron to the conduction band. Some other impurities, the so-called *acceptors* have the ability to extract an electron from the valence band, with a rather small expenditure of energy, allowing the remaining hole in the valence band to conduct. This can be done for example by raising the temperature. When the temperature is lowered again the electron will be returned to the valence band, or one can speak of the hole being bound by the acceptor. As in the case of the donors it can now be shown that the energy levels of the bound hole satisfy the equation

$$\left(-\frac{\hbar^2}{2m^{**}}\nabla^2 - \frac{e^2}{\varkappa r}\right) F = EF , \tag{1.9}$$

where m^{**} is an appropriate effective mass of the valence band.

2. – Band structures in silicon and germanium.

Before we can be more quantitative we must look somewhat more carefully at the effective masses. In silicon and germanium, in which impurity states have been studied in the greatest detail, the band structures are unfortunately fairly complex and we summarize here the relevant facts which have been obtained primarily from cyclotron resonance experiments [2].

2·1 *Conduction bands.* – In silicon the conduction band has 6 equivalent minima, located in the (0, 0, 1) and equivalent directions in k-space. Near the minimum in the (0, 0, 1)-direction the energy is given by

$$E(\boldsymbol{k}) = \frac{\hbar^2}{2m_l}(k_z)^2 + \frac{\hbar^2}{2m}(k_x^2 + k_y^2), \tag{2.1}$$

where k is reckoned from the position of the minimum and the effective masses have the values

$$m_l = 0.98\, m_0\,, \qquad m_t = 0.19\, m_0\,. \tag{2.2}$$

In germanium the minima are located in the (1, 1, 1) and equivalent directions, very likely at the edge of the Brillouin zone, in which case there are four of them. Near each minimum the energy is again given by expressions of the type (2.1) with

$$m_l = 1.60\, m_0\,, \qquad m_t = 0.0813\, m_0\,. \tag{2.3}$$

The surfaces of costant energy in these materials are evidently ellipsoids of revolution.

2·2 *Valence bands.* – In both silicon and germanium the valence bands have a single maximum at $k=0$. However this point is 4-fold degenerate, that is, four different bands come together and touch there, and the energy is determined as a function of k by the solution of a determinantal equation of the form

$$\left| D_{jj'}^{\alpha\beta} k_\alpha k_\beta - E\delta_{jj'} \right| = 0\,, \qquad jj' = 1\,, \ldots 4\,, \tag{2.4}$$

where the D's are certain known constants of the dimensions $\hbar^2/m_0$.

3. – Effective mass theory for impurity states in silicon and germanium.

In view of the somewhat complex band structures in silicon and germanium, the simple effective mass equation (1.2) must be modified [3].

For elliptical energy surfaces (donor states) the appropriate modification is

$$\left(E_j\left(\frac{1}{i}\,\boldsymbol{\nabla}\right) - \frac{e^2}{\varkappa r}\right) F_j = EF_j\,, \tag{3.1}$$

where $E_j(\boldsymbol{k})$ is the energy near the j'th minimum. A typical equation is

$$\left\{-\frac{\hbar^2}{2m_l}\frac{\partial^2}{\partial z^2} - \frac{\hbar^2}{2m_t}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) - \frac{e^2}{\varkappa r}\right\} F = EF\,. \tag{3.2}$$

The total wave functions of the impurity electron are then of the form

$$\psi = \sum_{j=1}^{N} \alpha_j F_j \varphi_j \,, \tag{3.3}$$

where N is the number of minima, φ_j is the Bloch wave at the j'th minimum and the α_j are numerical coefficients. In the ground state it is known in silicon and likely in germanium that all $|\alpha_j|$ are equal. Thus the wave-function is a superposition of wave-packets from each conduction band minimum.

For the degenerate valence band (acceptor states), the simple effective mass equation (1.2) must be replaced by the system of equations

$$\sum_{j'} \sum_{\alpha\beta} D_{jj'}^{\alpha\beta} \left(\frac{1}{i} \frac{\partial}{\partial x_\alpha}\right) \left(\frac{1}{i} \frac{\partial}{\partial x_\beta}\right) F_{j'} = \left(E - \frac{e^2}{\varkappa r}\right) F_j \qquad j, j' = 1, \ldots 4 \,, \tag{3.4}$$

and the total wave function of a bound hole is

$$\psi = \sum_{j=1}^{4} F_j \varphi_j \,, \tag{3.5}$$

where the φ_j are the four degenerate Bloch waves at the top of the valence band.

4. – Results for energy levels and comparison with experiment. Donor states.

When the effective mass equation (3.2) is solved by variational methods (with a precision of a few per cent), one obtains the following results.

In germanium the theoretical ground state energy is 0.0092 eV. Experimental values have been obtained for several donors from the temperature dependence of the Hall coefficient [4] and range from 0.0097 to 0.0127 eV. The agreement is fairly good. No experimental evidence for excited energy levels in germanium is as yet available.

For silicon the theory gives the following spectrum of energy levels (Table I).

TABLE I.

	Theoretical energies (in units of 0.01 eV)	Energy differences (in units of 0.01 eV)			
		Theory	P	As	Sb
$1s$	-2.9				
		1.8	3.45	4.21	3.18
$2p, m = 0$	-1.09				
		0.50	0.50	0·53	0.47
$2p, m = \pm 1$	-0.59				
		0.30	0.31	0.32	0.34
$3p, m = \pm 1$	-0.29				

This table shows also the result of infrared absorption experiments on P, As and Sb donors, in which the donor electrons were excited from the ground state to various p-like levels [5]. It is at once evident that the spacings between excited states are very similar for all three donors and in excellent agreement with the theory, while the position of the ground state varies substantially from one impurity to the next and in all cases is in serious disagreement with the theoretical prediction.

This state of affairs is not surprising. For while the $1s$ orbit of the lowest level penetrates into the immediate vicinity of the donor nucleus, where the potential certainly is not given by $-e^2/\varkappa r$, the p-like orbits of the excited states all have a node there and therefore are not appreciably affected by this region. For this reason the $1s$-level shows strong specific effects of each impurity, while the p-levels are well described by the effective mass equation with an $-e^2/\varkappa r$ potential.

4·1 *Acceptor States.* – In silicon, the ionization energies of Group III elements have been measured and vary from 0.45 eV for B to 0.16 eV for In [6]. This large variation shows that they cannot be adequately described by the effective mass theory. The excited acceptor states have been studied in beautiful measurements by HROSTOWSKI and KAISER [7], who have observed as many as 7 infra-red absorption lines for some acceptors. The theory suffers from a number of difficulties, among them the absence of accurate cyclotron resonance data, so that no conclusive comparison with experiment is yet possible.

The Group III acceptors in germanium have ionization energies in the range 0.0102 to 0.0112 eV, the theory gives 0.0089 eV in quite satisfactory agreement. Infrared absorption measurements on the excited states have not yet been reported.

One unambiguous result of the theory might be mentioned, namely that the acceptor ground state is 4-fold degenerate in both silicon and germanium. This conclusion has not yet been experimentally tested.

5. – Spin resonance experiments.

A few years ago Fletcher and co-workes observed spin resonance absorption by the donor electrons in silicon [7]. A sample containing bound donor electrons was placed in a static magnetic field H, and resonance absorption was observed for microwave frequencies given approximately by

$$\hbar\omega = 2\mu_0 H \,, \tag{5.1}$$

where μ_0 is the Bohr magneton. In fact Fletcher et al. found for each of several donors a number of absorption peaks, this number being in all cases equal

to $2I+1$, where I was the spin of the donor nucleus in question. This left no doubt that the multiplets were due to the hyperfine interaction of the electrons with the donor nuclei. The observed resonance frequencies were given by

$$\hbar\omega = 2\mu_0\left(H + \frac{8\pi}{3}\frac{\mu}{I} m_I |\psi(0)|^2\right), \tag{5.1'}$$

which differs from the simple expression (5.1) by the addition of the hyperfine interaction term. Here μ, I and m_I are respectively the magnetic moment, spin and magnetic quantum number of the donor nucleus and $|\psi(0)|^2$ is the probability density of the donor electron at the donor nucleus. A typical resonance pattern is shown in Fig. 2. From such measurements one obtains

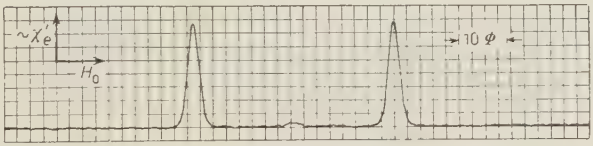


Fig. 2. – Electron spin resonance on phosphorus donors in silicon. $T=1.2$ °K; $\nu_e = 9\,000$ MHz.

the following experimental values of $|\psi(0)|^2$ (Table II), to which we have added the value produced by the effective mass theory.

TABLE II.

Donor	$\lvert\psi(0)\rvert^2$ (cm^{-3})
P	$0.44 \cdot 10^{24}$
As	$1.80 \cdot 10^{24}$
Sb	$1.20 \cdot 10^{24}$
Effective mass theory	$0.042 \cdot 10^{24}$

You will notice that the theoretical value is about one order of magnitude too small. We shall come back to this discrepancy in the following section.

6. – Corrections of the effective mass theory.

Let us focus our attention on a P-donor in silicon, the simplest one for our present purpose. This may be thought of as produced by starting from a perfect silicon lattice, adding a proton to one of the silicon nuclei, and adding an electron. Now in the immediate vicinity of the donor ion we expect the

effective mass equation to be substantially in error. For one thing, the effective additional potential is certainly not given by $-e^2/\varkappa r$, in fact for very small r it becomes $-e^2/r$. Furthermore, as a consequence of this rapid change of the potential, one finds that the whole concept of the effective mass does not apply to this region.

However, a lattice spacing or so away, the effective mass equation should be applicable but it must be solved with the *actual* binding energy of the donor electron which differs from the effective mass energy because of the complicated state of affairs in the neighborhood of the donor ion. This energy we take from experiment, -0.045 eV, instead of the effective mass value of -0.029 eV. When solving the effective mass equation with this energy and the boundary condition $F(\infty)=0$, the resulting function $F(\boldsymbol{r})$ tends to infinity as $r \to 0$. It is reasonable to consider this outside solution as valid up to the radius r_s of the Wigner Seitz sphere containing the donor ion. At this value of r, $|F(\boldsymbol{r})|^2$ is 9.8 times as large as when calculated by the effective mass theory. Taking this enhancement into account gives

$$|\psi(0)|^2_{\text{corrected}} = 9.8\,|\psi(0)|^2_{\text{eff. mass}} = 0.41 \cdot 10^{24}\ \text{cm}^{-3}\,, \tag{6.1}$$

in very good agreement with the experimental value of $0.44 \cdot 10^{24}\,\text{cm}^{-3}$ for phosphorus.

The values of $|\psi(0)|^2$ for other donors can also be roughly understood, if one allows for their different ionization energies and the different behavior of the donor electron in the cell of the donor ion.

7. – Many electron theory of impurity states.

In our discussions so far we have used a single electron wave function to describe the impurity states. For the excited states the results were so good that one suspects that the validity of the effective mass equation may be greater than the independent electron model from which it was derived. This is actually the case. One can prove the following theorem [10].

Consider a perfect semiconductor, with fixed nuclei, in its ground state. Now add an electron and a small additional positive point charge, q. Then, provided only that q is small enough, all low-lying energy levels of the entire many electron sistem are determined by an effective mass equation of the form

$$\left(-\frac{\hbar^2}{2m^*}\nabla^2 - \frac{e^2}{\varkappa r}\right) F = EF\,, \tag{7.1}$$

where m^* is an effective mass and $\varkappa$ is the static dielectric constant. All effects

of exchange and correlation are completely included in the two constants m^* and $\varkappa$. The same equation can also be rigorously derived for the energy levels of a hole.

The smallness of the charge q guarantees that all electron orbits are large compared to the lattice spacing, which is the basic prerequisite for the validity of the effective mass equation (7.1). This prerequisite is also satisfied when q has the finite value $|e|$, provided we restrict ouselves to highly excited orbits. This accounts for the good agreement, for such orbits, between theory and experiment.

8. – Miscellaneous experiments.

I should like to conclude by mentioning briefly a number of experiments which throw some light on the nature of the shallow impurity states.

8 1 *Double Resonance.* – In an electron spin resonance experiment the effective magnetic field acting on a donor electron is given by the parenthesis in Eq. (5.1′). It depends not only on the external magnetic field but also on the orientation of the donor nucleus.

Feher [11] has devised the following ingenious technique. In a fixed external field H, he varies the microwave frequency ω until he finds one of the resonances given by (5.1). He now holds this frequency fixed and increases the intensity of the signal until partial saturation occurs. In addition he now applies a radio-frequency signal of frequency ω_n, and varies ω_n until it equals one of the resonance frequencies of the donor nucleus, given by

$$\hbar\omega_n = 2\mu_{Si}\left(H \pm \frac{8\pi}{3}\mu_0 |\psi(0)|^2\right) \tag{8.1}$$

(cf. Eq. (5.1′). At such a frequency the magnitude of the *electron* resonance absorption, A, is altered.

By measuring A as a function of ω_n one can therefore again obtain the value of $|\psi(0)|^2$, if the moment of the donor is known. Conversely, the ratio of the effective moments of two isotopes may be very accurately determined by this technique [12].

Actually, the electron spin interacts not only with the donor nucleus but also (more weakly) with the magnetic moments of the ^{29}Si nuclei within its orbit. Whenever ω_n equals the resonance frequency of one of these moments, A is altered. In this way the values of all $|\psi(\boldsymbol{r}_l)|^2$ are obtained, where $\boldsymbol{r}_l$ is any lattice site.

8·2 *Strong magnetic fields.* – Since in some semiconductors the effective masses are very small, external magnetic fields have a strong effect on the orbital motion of impurity electrons [13]. Thus their ionization energy is appreciably increased when the parameter

$$\gamma = \frac{eh}{2m^*c} H/E_0 , \tag{8.2}$$

becomes comparable with 1. (E_0 is the ionization evergy in the absence of a magnetic field). In silicon and germanium fields of the order of 10^5 to 10^6 oersted are required to see appreciable effects, and no such measurements have yet been reported. But in InSb, where $m^*/m_0 = 0.013$, the increase of ionization energy has been established by KEYES and SLADEK [14], working with fields of the order of 1000 Oe.

8·3 *Stark effect.* – Because of their large orbits, the perturbing potential due to an external field $\mathcal{E}$,

$$\delta V = - e\mathcal{E}z , \tag{8.3}$$

should have a large effect on impurity states, but no measurements have as yet been reported.

The first order Stark effect vanishes when calculated in the effective mass approximation. The second order energy shift is given, approximately by

$$\delta_2 E \approx \frac{(e\mathcal{E}R)^2}{\Delta E} ,$$

where R is the radius of the orbit and ΔE is an average energy denominator. Substituting typical numbers for a highly excited impurity state in silicon ($R = 10^{-6}$ cm, $\Delta E = 5 \cdot 10^{-3}$ cm), gives an energy shift of the order of $2 \cdot 10^{-3}$ eV, comparable to the total energy of such a level.

REFERENCES

[1] W. KOHN: *Shallow Impurity States in Silicon and Germanium*, to be published in Vol. **5** of *Solid State Physics*, Editors F. SEITZ and D. TURNBULL (New York).

[2] G. DRESSELHAUS, A. F. KIP and C. KITTEL: *Phys. Rev.*, **98**, 368 (1955).

[3] C. KITTEL and A. H. MITCHELL: *Phys. Rev.*, **96**, 1488 (1954); J. M. LUTTINGER and W. KOHN: *Phys. Rev.*, **97**, 869 (1955); W. KOHN and J. M. LUTTINGER: *Phys. Rev.*, **98**, 915 (1955).

[4] T. H. Geballe and F. J. Morin: *Phys. Rev.*, **95**, 1085 (1954).
[5] G. S. Picus, E. Burstein and S. Henvis: *Journ. Phys. Chem. Solids*, **1**, 75 (1956). In order to compare the theory with these experiments, the $3p$, $m=0$ and $2s$ levels for which the theory indicates very weak absorption have been omitted.
[6] Ref. [4] and F. J. Morin, J. P. Maita, R. G. Shulman and N. B. Hannay: *Phys. Rev.*, **96**, 833 (1953).
[7] To be published; see also E. Burstein, G. Picus, S. Henvis and Wallis: *Journ. Phys. Chem. Solids*, **1**, 65 (1956).
[8] R. C. Fletcher W. A. Yager, G. L. Pearson, A. N. Holden, W. T. Read and F. R. Merritt: *Phys. Rev.*, **94**, 1392 (1954); R. C. Fletcher, W. A. Yager, G. L. Pearson and F. R. Merrit: *Phys. Rev.*, **95**, 844 (1954).
[9] F. J. Morin *et al.*: ref. [6].
[10] W. Kohn: *Phys. Rev.*, **105**, 509 (1957).
[11] G. Feher: *Phys. Rev.*, **103**, 834 (1956) and private communication.
[12] G. Feher: to be published in *Phys. Rev.*
[13] Y. Yafet, R. W. Keyes and E. N. Adams: *Journ. Phys. Chem. Solids*, **1**, 137 (1956).
[14] R. W. Keyes and R. J. Sladek: *Journ. Phys. Chem. Solids*, **1**, 143 (1956).

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Recombination Processes in Semiconductors.

P. AIGRAIN

Laboratoire de Physique, École Normale Supérieure - Paris

1. – Introduction.

The recombination of minority carriers with majority carriers in a semiconductor, or the inverse process of hole-electron pair production is a problem of great physical and technological interest. Indeed the recombination rate of excess minority carriers is one of the most important properties of a semiconducting crystal for device use.

This recombination rate is often specified in terms of « lifetime for recombination » τ, i.e. the mean time a minority carrier spends in the crystal as a free carrier before recombining. The words « as a free carrier » are of importance as a minority carrier, before recombining, may very well get « trapped », then later released, on crystal defects. These trapping phenomena, when present, complicate seriously the measurement of τ since a minority carrier may spend a time much larger then τ in traps before recombination. To make certain that the time measured is τ one is led to measure, instead of a time, the length a minority carrier travels before recombination or « diffusion length » L. The quantities L and τ are related by

$$L = \sqrt{D\tau} \,. \tag{1}$$

where D is the diffusion constant of the minority carrier.

Methods for measuring either lifetime or diffusion length are numerous and their study would be outside the scope of this lecture. But one must realize that only those methods which are basically diffusion length measurements can be considered perfectly reliable in the presence of trapping phenomena.

It has long been recognized that the recombination rate of minority carriers near the surface of a semiconductor was much larger than in the bulk. Thus one is led to study separately bulk and surface recombination.

2. – Bulk recombination.

Recombination in the bulk may occur both through intrinsic processes (independent of the presence of a few crystal defects) and extrinsic processes (recombination catalysed by defects).

2·1. *Radiative recombination.* – Of the intrinsic processes the most easily understood is radiative recombination whereby the gap energy is carried away by a photon. This is the inverse process of hole-electron pair production by photons, which is responsible for the absorption edge. Thus SHOCKLEY and V. ROOSBROECK showed that recombination rate through this process could be computed by applying the principle of detailed balancing, once one knows the absorption constant versus wavelength.

It can thus be shown that in a high purity crystal of germanium or silicon at low injection levels, this process is completely unimportant in controlling lifetime. In germanium the radiative lifetime is about 0.750 s at room temperature, while the observed lifetimes range only up to 0.005 s. However, radiative recombination is important in semiconductors with low gap, such as InSb, where it might be the most important process. Furthermore radiative recombination is a bimolecular process, so that it increases with injection level. This is well shown in recent experiments by BENOIT in which, at extremely high injected densities of injected carriers, radiative recombination took over as the main recombination process.

It should be remarked that band to band recombination in a semiconductor is usually a first order forbidden process unless the valence and conduction band extrema lie at the same point in K space (usually $K = (000)$). This is not the case for Ge so that recombination occurs through a second order process involving the emission of a photon and the simultaneous emission or absorption of a phonon. But the presence of a high lying conduction band at $K = (000)$ (in Ge such a band exists only 0.10 eV above the conduction band minima) can act as a catalyst: any electron scattered into the (000) band recombines very quickly with a hole. At room temperature in Ge, 65 % or radiative recombination occurs through the (000) band. As a result band to band recombination light is often strongly reabsorbed, with the production of a hole electron pair: instead of a true recombination one has a sudden transfer of excitation. In the limit where the recombination light is very strongly reabsorbed one has a diffusion of the excitation.

2·2. *Auger processes.* – PINCHERLE has studied theoretically the possibility of a recombination process whereby the excitation energy is transferred to a third free particle: this is the equivalent of an inverse Auger process. Such a recombination process depends critically on the exact band shape: indeed it is easily seen that crystal momentum and energy are usually hard to conserve simultaneously.

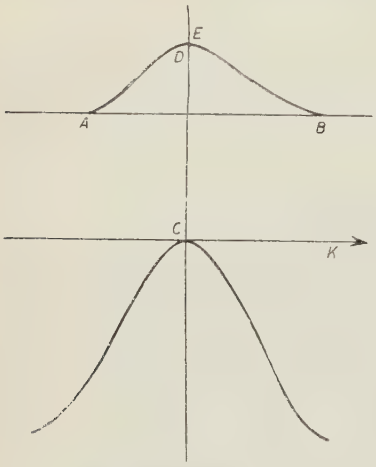


Fig. 1. – Hypothetical profile of energy bands in a semiconductor.

One case where Auger recombination may be operating is that of Tellurium. DE CARVALHO, using a PME technique showed that τ in Te is structure insensitive and varies very fast with temperature (from 50 μs at 77 °K to 10^{-8} s at room temperature). In fact τ varies exponentially, with an activation energy which is precisely the band gap of Te (0.34 eV). An explanation can be found if recombination is through an Auger process and if the third particle must be an electron (minority carrier). Why one type of carrier may be much more efficient in this than the other, can be explained with the help of the hypothetical band shape (Fig. 1).

A process in which two electrons in A and B collide with a hole in C, thus exciting an electron in D while the other electron and the hole recombine, conserves both energy and momentum. This also shows that Auger recombination can only be expected to be effective when some accidental near coincidence in gap energy occurs.

2·3. *Defect catalyzed bulk recombination.* – This is the most important recombination process in crystals like Ge, Si etc.. SHOCKLEY, READ and HALL have given a phenomenological theory of this type of recombination assuming that the defect produces a deep lying level in the forbidden band and recombination is a two step process, whereby a free particle gets trapped and then recombines with a carrier of opposite polarity. They show that

$$\tau = \tau_{n_0} \frac{n_0 + n_1}{n_0 + p_0} + \tau_{p_0} \frac{p_0 + p_1}{n_0 + p_0}, \tag{2}$$

where n_0 and p_0 are the electron and hole densities at equilibrium, and n_1 and p_1 the electron and hole densities in a sample in which the Fermi level would lie at the trap level. Formulae for the recombination rate far away from equilibrium ($n \gg n_0$, $p \gg p_0$) are also available. τ_{n_0}, τ_{p_0} are parameters,

related to the trap capture cross-sections σ_n for electrons and σ_p for holes. (Usually $\tau_{n_0} \gg \tau_{p_0}$, because $\sigma_n \ll \sigma_p$).

Although this theory can explain most of the bulk recombination processes one still has to answer the following questions:

A) What are the traps?

B) Where does the trap level lie?

C) How is the trapping energy dissipated?

Answers to the first question can be found by introducing intentionally some defects, or studying the correlation between the presence of defects and high recombination rates. Thus for Ge dislocations, vacancies, copper, nickel were found to be efficient recombination centers.

The second question can be answered by studying the temperature dependence of τ (but one has to make assumptions on the temperature dependence of τ_{n_0}, τ_{p_0}) or the dependence of τ on the Fermi-level, which can be varied by doping (but one has to maintain the trap density constant at the same time). When the trapping energy is dissipated as light, this quantity is obtained directly from spectroscopic measurements. A rather curious experimental fact is that, in germanium, most of the recombination centers studied give levels 0.15 to 0.25 eV above the valence band. As concerns the third question, light emission from recombination through traps at liquid air temperature has been picked up at 2.35 μm (0.52 eV) by BENOIT, and later by NEWMANN. NEWMANN has showed that the traps involved were dislocations. Copper also produces radiative recombination at liquid air temperature, at 0.47 eV. But at room temperature no light is observed, thus showing that recombination is usually a non-radiative process, in which the energy is carried by numerous phonons. The temperature variation of radiative efficiency, measured by BENOIT for dislocations agrees well with the idea of a competition between a temperature independent radiative process, and a non-radiative process having a cross-section which varies exponentially with temperature (activation energy 0.14 eV) as expected from the theory of multiphonon processes.

In silicon HAYNES has shown that donor and acceptor centers can act as rather inefficient recombination centers, and the trapping is radiative.

3. – Surface recombination.

It has long been known that the surface of a semiconductor was a region of high recombination. Surface recombination is usually specified by the surface recombination velocity S, which being the ratio of recombinations per unit time per unit area to the density of excess minority carriers has the

dimensions of a velocity. The high values of S (from 20 to 10^6 cm/s) are of course to be associated with a high surface density of recombination centers, and a theory of the Shockley-Read type should apply. Complications occur, however, because the density of surface traps is high enough for a barrier to be formed at the free surface, so that hole and electron densities at the surface differ, from those in the bulk, even at equilibrium.

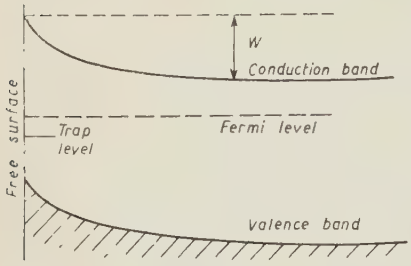


Fig. 2.
Energy barrier at a free surface.

In the case shown in Fig. 2 the electron density will be increased by the factor $\exp[-w/kT]$, and the hole density increases correspondingly. Furthermore, for those semiconductors of most interest (Ge, Si, SbIn etc.) the free surface is covered by an oxide layer which may itself contain charged traps, the presence of which modifies the barrier height. Thus S will depend not only on the nature and density of surface recombination centers, but also on previous history, atmosphere etc., which control the charge per unit area in the oxide layer, and thus the barrier height.

The modification of the Shockley-Read theory for treating surface recombination is due to BARDEEN and SHOCKLEY. It introduces barrier height W as a parameter.

A further complication, about which little has been done is connected with the fact that a one electron theory is actually not expected to hold too well in a space charge layer. For example if a negatively charged surface trap releases an electron thermally, to be conduction band at the surface, the energy of the system is not changed by the amount E_T (Fig. 2) but by a smaller amount, since making one trap neutral also reduces the mutual repulsion energy of the charged surface traps. In the bulk free electrons shield out this interaction, but this cannot occur in a space charge layer. This point is not properly taken into account in existing theories.

Assuming the Bardeen-Shockley theory to hold, it is obvious that surface recombination can be studied more easily than bulk recombination, because one can, by applying a transverse field vary the barrier height and thus the parameters n_0, p_0 (eq. (2)) without changing the trap density. MANY has carried out such experiments which show that the kind of behaviour predicted by eq. (2) is obeyed. It is hard to know whether the trap levels E_T given by these experiments are to be trusted in view of the objection of the preceeding Section.

At any rate MANY's results show that surface recombination centers, like those in the bulk, are usually acceptors like (high cross-section for capture of holes $\sigma_n/\sigma_p = 10^{-2}$) on germanium.

Nothing is known about the nature of these traps, the density of which (10^{10} to 10^{12} cm^{-2} in general) excludes any possibility that dislocation exit points might be involved. The traps may be region of mismatch between the semiconductor and the oxide layer.

Attempts to observe light emission from surface recombination have failed to date, but the possibility that the transition may be partially radiative cannot be ruled out.

4. – Conclusion.

This short review of recombination processes in semiconductors shows that one is still very far from having a complete understanding of these processes. Furthermore nearly all the experimental work has been on Germanium, with Silicon a poor second. The problems of surface recombination may be expecially difficult because a one electron picture is not expected to hold.

SUPPLEMENTO AL VOLUME VII, SERIE X
DEL NUOVO CIMENTO

N. 2, 1958
1° Trimestre

Halbleiter mit kleiner scheinbarer Masse der Ladungsträger.

O. MADELUNG

Forschungslaboratorium der Siemens-Schuckert-Werke AG - Erlangen

Unter den charakteristischen Parametern eines Halbleiters nimmt die *scheinbare Masse* m^* der Ladungsträger eine wichtige Stellung ein. Es sind im wesentlichen drei Stellen, an denen sie in den Formalismus der Halbleitertheorie eingeht. Einmal ist sie mitbestimmed für die *Beweglichkeit* μ der Ladungsträger, also für deren Verhalten unter dem Einfluß äußerer Felder durch die Gleichung

$$\mu \approx e\tau/m^* , \tag{1}$$

wo τ die Relaxationszeit ist, die ebenfalls von m^* abhängt. Für die durch die Wechselwirkung der Ladungsträger mit dem Gitter begrenzte Beweglichkeit gilt beispielsweise $\mu \sim m^{*-\frac{5}{2}}$. Weiter ist die Beschleunigung, die ein Ladungsträger in einem *Magnetfeld* B erfährt, proportional der Cyclotron-Resonanz-Frequenz

$$\omega_c = eB/m^* . \tag{2}$$

Alle magnetischen Effekte hängen also quantitativ von der Größe der scheinbaren Masse ab. Insbesondere tritt in allen galvanomagnetischen und thermomagnetischen Prozessen immer das Magnetfeld in der Kombination $\omega_c\tau \approx \mu B$ auf. Diese Effekte sind also besonders groß in Halbleitern kleiner scheinbarer Masse bzw. großer Beweglichkeit der Ladungsträger. Schließlich ist die Beziehung zwischen Energie und Wellenzahl im *Bändermodell* in der Nähe der Bandränder in erster Näherung gegeben durch

$$E = \hbar^2 k^2/2m^* . \tag{3}$$

Hier bestimmt also die scheinbare Masse die « Krümmung » der $E(k)$-Abhängigkeit und damit die Energieverteilung der in dem betreffenden Band befindlichen Ladungsträger.

Der Begriff einer konstanten skalaren scheinbaren Masse ist jedoch nur in erster Näherung im Bändermodell des isotropen Halbleiters anwendbar. Werden in einem Band auch Energieterme mit Ladungsträgern besetzt, die nicht mehr durch die Approximation (3) beschrieben werden können, so wird die mittlere scheinbare Masse der Ladungsträger-Gesamtheit abhängig von der Besetzung des Bandes, also etwa von der Ladungsträgerdichte bzw. der Temperatur. Bei Berücksichtigung der Anisotropie eines Halbleiters wird schließlich $1/m^*$ ein Tensor. Bei der Untersuchung eines Halbleiters kann man jedoch immer zunächst zur Gewinnung allgemeiner Informationen über seine Eigenschaften mit einer konstanten skalaren scheinbaren Masse rechnen.

Die numerischen Werte der scheinbaren Massen liegen im allgemeinen bei Halbleitern zwischen 0.1 und 1 m_0, wo m_0 die Elektronenmasse ist. Bei einigen Halbleitern (Mg_2Sn, Mg_2Sb_3 u.a.) treten auch Werte $> m_0$ auf. Scheinbare Massen $< 0.1\ m_0$ sind bisher nur in Ge, HgSe, HgTe, InSb und InAs gemessen worden. In Ge ist lediglich die transversale Masse der Elektronen $< 0.1\ m_0$ ($m_{n_t} = 0.082\ m_0$), sowie die scheinbare Masse der Löcher des zweiten Teilbandes des Valenzbandes ($m_{p_2} = 0.044\ m_0$). Hier liefert jedoch die wesentlich größere longitudinale Masse der Elektronen von $1.58\ m_0$ ein mittleres $m_n = 0.22\ m_0$, während im Valenzband die Dichte der «leichten» Löcher nur wenige Prozent der gesamten Löcherdichte beträgt. In Ge sind also keine Besonderheiten durch kleine scheinbare Massen zu erwarten. Während für HgSe und HgTe noch keine genauen Werte vorliegen ($m_n < 0.1\ m_0$, $\mu_n > 10\,000$ cm²/Vs), sind die Werte bei InSb und InAs recht genau bekannt. Man findet als mittlere scheinbare Massen der Elektronen in InSb den Wert $0.037\ m_0$ und in InAs $0.064\ m_0$, aber in beiden Fällen $m_p > 0.1\ m_0$. Wir werden uns im Folgenden mit diesem beiden Halbleitern beschäftigen.

InSb und InAs gehören zu den sogenannten III-V-Verbindungen, die zuerst von H. Welker [1] untersucht wurden. Diese Halbleiter sind Verbindungen zwischen Elementen der dritten Gruppe und der fünften Gruppe des periodischen Systems der Elemente, insbesondere zwischen B, Al, Ga, In einerseits und N, P, As, Sb andererseits. Sie kristallisieren in der Zinkblendestruktur, die in die Diamantstruktur der halbleitenden Elemente Diamant, Si, Ge und α-Sn übergeht, wenn man die ungleichartigen Atome der Verbindung durch gleichartige ersetzt. In diesen Strukturen ist jedes Atom von vier nächsten Nachbarn umgeben und die chemische Bindung erfolgt durch abgesättigte Elektronenpaare zwischen je zwei nächsten Nachbarn. Aus dieser Ähnlichkeit des Bindungsmechanismus erklärt sich die enge Verwandtschaft der III-V-Verbindungen zu den halbleitenden Elementen der IV. Gruppe. Während in den elementaren Halbleitern die chemische Bindung weitgehend homöopolar ist, macht sich beim Übergang zu den III-V-Verbindungen ein (durch die ungleiche Elektronegativität benachbarter Atome bedingter) heteropolarer Bindungsanteil bemerkbar, der sich beim Fortschreiten in der «isoelektro-

nischen» Reihe IV-IV, III-V, II-VI, I-VII verstärkt. Eine quantitative Theorie der chemischen Bindung in den III-V-Verbindungen und damit eine exakte Erklärung der Unterschiede zwischen den Eigenschaften der Elemente der IV. Gruppe und den III-V-Verbindungen steht noch aus. Kennzeichnend ist vor allem ein Anwachsen der Breite der verbotenen Zone und der Elektronenbeweglichkeit, sowie ein Abfall der scheinbaren Massen der Elektronen beim Übergang von den Elementen zu den III-V-Verbindungen. Theoretische Ansätze zur Erklärung dieser letzteren Tatsache blieben bisher erfolglos [2, 3]. Lediglich Gubanow [3] konnte zeigen, daß selbst bei konstant bleibender scheinbarer Masse mit wachsendem heteropolarem Bindungsanteil die Beweglichkeiten anwachsen können.

Wir betrachten nun die Auswirkungen der kleinen scheinbaren Elektronenmassen in InSb und InAs auf deren physikalische Eigenschaften. Als erstes zu erwähnen sind die mit den kleinen scheinbaren Massen verbundenen extrem hohen Elektronenbeweglichkeiten von 77000 cm^2/Vs in InSb und 27000 cm^2/Vs in InAs. Da die Löcherbeweglichkeit in beiden Halbleitern relativ klein ist (InSb: etwa 1000 cm^2/Vs, InAs: 280 cm^2/Vs), ist neben der hohen Elektronenbeweglichkeit ein großes Beweglichkeitsverhältnis charakteristisch. Dies hat z.B. zur Folge, daß bei einer gegebenen Temperatur der Halbleiter nicht im Zustand der Eigenleitung seinen höchsten Widerstand besitzt, sondern im Gebiet der vorwiegenden p-Leitung, genauer bei $p/n = \mu_n/\mu_p$. Messungen der Temperaturabhängigkeit der Leitfähigkeit zeigen also bei p-leitenden Proben ein Unterschreiten der Eigenleitfähigkeits-Geraden vor Eintritt in die Eigenleitung [4]. Entsprechend können p-leitende Proben bei gegebener Temperatur einen höheren Wert des Hall-Koeffizienten annehmen, als eigenleitende Proben.

Wichtiger als diese Anomalien sind die Auswirkungen der hohen Beweglichkeit auf die galvanomagnetischen Effekte. Dies gilt vor allem für die Widerstandsänderung im Magnetfeld, die ja proportional dem Quadrat des Produktes $\mu \cdot B$ ist. Sie wird hervorgerufen durch die Tatsache, daß die Ladungsträger im Halbleiter keine einheitliche Geschwindigkeit besitzen. Da die zur Kompensation der auf die Ladungsträger wirkenden Lorentz-Kraft im Halbleiterinneren auftretende Hall-Feldstärke nur die Ablenkung der Ladungsträger einer mittleren Geschwindigkeit kompensieren kann, werden Ladungsträger mit einer höheren Geschwindigkeit doch durch das Magnetfeld abgelenkt, während die Ablenkung langsamerer Teilchen durch das Hall-Feld überkompensiert wird. Dies führt zu einer Verlängerung der Bahn der Ladungsträger und damit zu einer Widerstandserhöhung. Bei Halbleitern hoher Beweglichkeit tritt aber noch ein zusätzlicher Effekt auf. Bei der üblichen Meßanordnung des Hall-Effektes werden dünne Stäbe benutzt und die Hall-Spannung tritt auf, um die Ladungsträger trotz der Einwirkung der Lorentz-Kraft zu zwingen, in Richtung der Stabachse zu fließen. Wir betrachten jetzt dagegen den Fall,

daß der Halbleiter die Form einer breiten Platte besitzt und legen speziell das Magnetfeld senkrecht zur Platte in die z-Richtung und das primäre elektrische Feld in die x-Richtung. Dann tritt keine Hall-Spannung auf, da die Elektronen jetzt ungehindert in die y-Richtung abgelenkt werden können und unter einem Winkel ϑ zur x-Richtung fließen.

Dies ergibt aber wiederum eine Bahnverlängerung und somit eine Widerstandserhöhung ϱ_H/ϱ_0, die gegenüber der Widerstandserhöhung des oben betrachteten Falles um den Faktor $(1+\mathrm{tg}^2\,\vartheta)$ größer ist. Für $\mathrm{tg}\,\vartheta$ ergibt die Theorie in erster Näherung den Wert $10^{-8}\mu B$, wenn μ in $\mathrm{cm}^2/\mathrm{Vs}$ und B in Gauß gemessen wird. Man erkennt sofort, daß die Änderung der Geometrie der Probe keinen wesentlichen Einfluß auf die Widerstandsänderung hat, solange μB klein gegen 10^8 ist, also in praktischen Fällen, bei welchen Magnetfelder der Größenordnung 10^4 Gauß benutzt werden, solange μ kleiner als 10 000 $\mathrm{cm}^2/\mathrm{Vs}$ ist. Dieser Wert wird aber von InSb und InAs weit überschritten.

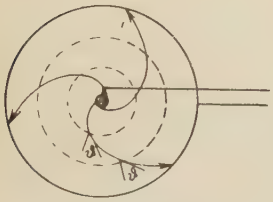


Fig. 1. – Corbino-Scheibe.

Hier findet man im Grenzfall eines unendlich breiten Präparates Widerstandsänderungen, die die übliche Widerstandsänderung weit übertreffen. Dieser Grenzfall läßt sich leicht realisieren durch die sogenannte *Corbino-Scheibe* (Fig. 1). Benutzt man nämlich eine kreisförmige Scheibe, deren Zentrum die eine Elektrode und deren Peripherie die andere Elektrode bildet, so kann sich ebenfalls keine Hall-Spannung ausbilden. Die Äquipotentiallinien des elektrischen Feldes bleiben Kreise und die Ladungsträger laufen auf spiralförmigen Bahnen. Fig. 2 zeigt die Geometrieabhängigkeit der Widerstandsänderung an vier verschieden geformten Proben aus InSb sonst gleicher Parameter (μ, n usw.).

Eine hohe Elektronenbeweglichkeit ist jedoch nicht nur zur Erzielung einer großen Widerstandsänderung wichtig. Die Hall-Spannung selbst wächst ja auch mit zunehmender Beweglichkeit der Ladungsträger. Während bei den meisten Festkörpern die Hall-Spannung nur mit empfindlichen Galvanometern gemessen werden kann, ist bei InSb und InAs V_H so groß, daß der üblichen Anordnung zur Messung des Hall-Effektes eine Hall-Leistung entnommen werden kann (*Hall-Generator*). Der Wirkungsgrad einer solchen Anordnung ist der Quotient aus Hall-Leistung ($L_H \sim V_H^2\,\sigma$, wir vernachlässigen im weiteren alle Geometrie- und Anpassungsfaktoren) und Primärleitung ($L_{pr} \sim i_{pr}/\sigma$), also

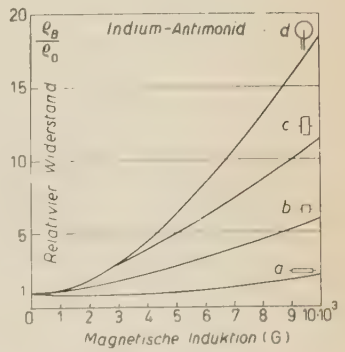


Fig. 2. – Relative Widerstandsänderung von vier InSb-Proben verschiedener geometrischer Form.

wegen $V_H \sim R i_{pr} B_z$ ($R =$ Hall-Koeffizient $\sim \mu/\sigma$): Wirkungsgrad $\sim (\mu B_z)^2$. Der Wirkungsgrad des Hall-Generators ist also bei InSb um mehr als zwei Zehnerpotenzen größer als bei Germanium. Auch die weiteren Forderungen, die zur technischen Anwendung an den Hall-Generator zu stellen sind (hoher spezifischer Widerstand d.h. hoher Hall-Koeffizient aus Anpassungsgründen, geringe Temperaturabhängigkeit des Hall-Koeffizienten), werden von diesen zwei Halbleitern erfüllt. Die technische Anwendung erstreckt sich vornehmlich auf:

1) Messung von Magnetfeldern bei konstant gehaltenem Primärstrom. Hier konnte ein Gerät entwickelt werden, daß geeignet ist, Magnetfelder bis herab zu 10^{-5} Gauß zu messen [5].

2) Multiplikation zweier elektrischer Größen, die proportional zum Magnetfeld und zum Primärstrom sind. Beispiel: Leistungsmessung, Messung des Drehmoments eines Elektromotors.

Neben den hohen galvanomagnetischen Effekten treten dank der großen Beweglichkeiten auch hohe thermomagnetische Effekte, wie der Nernst-Effekt auf.

Bei den galvanomagnetischen Effekten in InSb und InAs sind noch drei Anomalien zu erwähnen, deren Deutung bisher aussteht. Dies ist einmal eine doppelte Vorzeichenumkehr des Hall-Koeffizienten von InAs-Proben im Gebiet der vorwiegenden p-Leitung [6], eine im Gegensatz zur Theorie viel zu kleine magnetische Widerstandsänderung in stabförmigen InAs-Präparaten [7] und das Auftreten einer negativen Widerstandsänderung im Magnetfeld in InSb bei tiefen Temperaturen [8]. Die Klärung der beiden erstgenannten Anomalien hängt sicher mit dem Problem der Bandstruktur des InAs zusammen, die noch nicht in dem gleichen Masse bekannt ist, wie etwa die des Germaniums oder des Siliziums. Zur Klärung der negativen Widerstandsänderung liegen bereits theoretische Ansätze vor, doch steht eine quantitative Theorie noch aus [9].

Ein weiterer Effekt, der zuerst an InSb gefunden wurde, ist der sogenannte *magneto-optische* Effekt. Er beruht auf der Tatsache, daß in einem Magnetfeld, die quasikontinuierlichen Terme eines Energiebandes sich zu diskreten eindimensionalen Teilbändern zusammenschließen. Liegt das Magnetfeld in der z-Richtung, so haben diese Teilbänder die Form

$$E_n = \frac{\hbar^2 k_z^2}{2m^*} + \hbar\omega_c \left(n + \frac{1}{2}\right). \tag{4}$$

Hier ist ω_c wieder die Cyclotron-Resonanzfrequenz eB_z/m^*. Man erkennt aus (4) sofort, daß das unterste Teilband gegenüber dem magnetfeldfreien Fall um $\Delta E_L = \hbar\omega_c/2$ angehoben wird, die Breite der verbotenen Zone also um

ΔE_L + einer entsprechenden Absenkung des Valenzbandes ΔE_V wächst. Dieser Energiezuwachs und die Aufspaltung der Teilbänder ist umso größer, je kleiner die scheinbare Masse der Ladungsträger ist. Fig. 3 und 4 zeigen die experimentelle Bestätigung dieses Effektes. In Fig. 3 [10] ist die Verschiebung der Absorptionskante (also die Verbreiterung der verbotenen Zone) von InSb im Magnetfeld gezeigt. Die Größe der Verschiebung entspricht angenähert der Theorie. Fig. 4 zeigt die relative Durchlässigkeit von InSb bei Zimmertemperatur aufgetragen gegen die Energie der eingestrahlten Phononen für verschiedenes Magnetfeld [11]. Charakteristisch sind die drei Durchlässigkeitsminima, deren Lage magnetfeldabhängig ist. Sie werden optischen Übergängen von verschiedenen magnetischen Teilbändern des Valenzbandes in verschiedene Teilbänder des Leitungsbandes zugeschrieben. Mit den bekannten Werten der scheinbaren Massen lassen sich diese Minima deuten als Übergänge zwischen Teilbändern mit den Quantenzahlen $0 \rightarrow 0$, $1 \rightarrow 0$, $0 \rightarrow 1$.

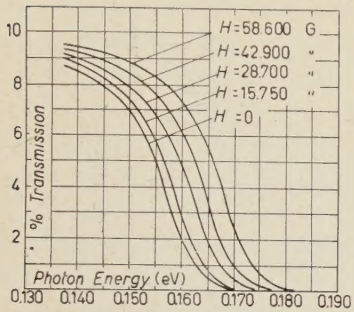


Fig. 3. – Absorptionskante von InSb für verschiedene Magnetfelder.

Mit dem Auftreten der scheinbaren Masse in Gl. (3), also in der Abhängigkeit der Energie von der Wellenzahl sind zwei weitere Anomalien verknüpft. Kleine scheinbare Masse bedeutet in (3) eine starke Krümmung der $E(k)$-Abhängigkeit und damit eine geringe Zustandsdichte am Bandrand. Mit wachsender Elektronenkonzentration im Leitungsband steigt also die Energie des jeweils tiefsten besetzbaren Terms wesentlich schneller an, als dies bei Halbleitern mit größerer scheinbarer Elektronenmasse der Fall ist. Dies hat z.B. zur Folge, daß die Absorptionskante, deren Lage dem energetischen Abstand zwischen dem höchsten mit Elektronen besetzten Term des Valenzbandes und dem tiefsten freien Term des Leitungsbandes entspricht, sich mit wachsender Elektronenkonzentration, also auch mit wachsendem Donatorengehalt nach kürzeren Wellenlängen verschiebt [12]. Weiter hat die geringe Zustandsdichte am Rande des Leitungsbandes zur Folge, daß mit wachsender Besetzung des Bandes auch Gebiete erreicht werden, in denen die Approximation (3) nicht mehr

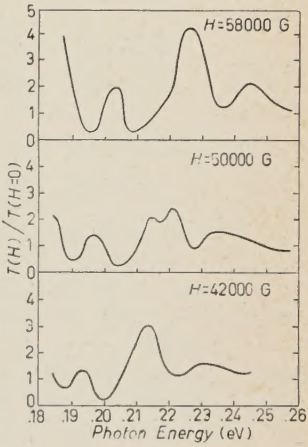


Fig. 4. – Relative Durchlässigkeit von InSb bei Zimmertemperatur in Abhängigkeit von der Energie der eingestrahlten Phononen bei verschiedenem Magnetfeld.

zulässig ist. Die mittlere scheinbare Masse der Elektronen wird dann abhängig von der Elektronendichte selbst bzw. von die Elektronendichte bestimmenden Faktoren, wie Donatorenkonzentration oder Temperatur. Als ein Beispiel zeigt Fig. 5 eine Zusammenstellung von experimentellen Ergebnissen verschiedener Autoren über die scheinbare Elektronenmasse in InSb. Die Punkte sind Ergebnisse von Cyclotronresonanz-Messungen, die ausgezogenen Linien stammen aus Thermokraftmessungen, während die gestrichelten Linien aus indirekten Verfahren herrühren (Suszeptibilitätsmessungen, Beweglichkeitsmessung im Bereich der Ionenstreuung u.a.). Ohne auf eine genaue Diskussion der einzelnen Kurven einzugehen, läßt sich aus der Figur bereits deutlich erkennen, daß die scheinbare Masse mit wachsender Temperatur, also wachsender Besetzung des Leitungsbandes mit Elektronen ansteigt.

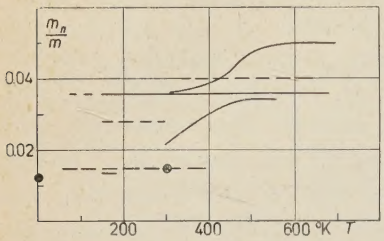


Fig. 5. – Zusammenstellung verschiedener experimentell gewonnener Werte der scheinbaren Elektronenmasse in InSb.

LITERATURVERZEICHNIS

[1] H. Welker: *Zeits. Naturforsch.*, **7***a*, 744 (1952).
[2] I. Adawi: *Phys. Rev.*, **105**, 789 (1957).
[3] A. I. Gubanow: *Journ. Techn. Phys. USSR*, **26**, 2170 (1956).
[4] O. G. Folberth und O. Madelung: *Zeits. Naturforsch.*, **8***a*, 673 (1953).
[5] H. Hieronymus und H. Weiss: *Siemens-Zeitschrift* **31**, 404 (1957).
[6] O. G. Folberth und H. Weiss: *Zeits. Naturforsch.*, **11***a*, 510 (1956).
[7] H. Weiss: *Zeits. Naturforsch.*, **12***a*, 80 (1957).
[8] H. Fritzsche und K. Lark-Horovitz: *Phys. Rev.*, **99**, 400 (1955).
[9] Vgl. etwa I. M. Mackintosh: *Proc. Phys. Soc.*, **69** B, 403 (1956); C. Rigeaux und J. M. Thuillier: *Compt. Rend. Acad. Sci.*, **242**, 2710 (1956).
[10] E. Burstein, G. S. Picus, H. A. Gebbie und F. Blatt: *Phys. Rev.*, **103**, 826 (1956).
[11] E. Burstein und G. S. Picus: *Phys. Rev.*, **105**, 1123 (1957).
[12] E. Burstein: *Phys. Rev.*, **93**, 632 (1954).

INDICE DEL SUPPLEMENTO

AL VOLUME VII, SERIE X, DEL
NUOVO CIMENTO
Anno 1958

H. E. Moses – A spinor representation of Maxwell's equations . . . . . pag. 1

V. S. Barašenkov and B. M. Barbašev – Statistical weight of a system of particles with arbitrary spins . . . . . » 19

L. Pàl – On the theory of stochastic processes in nuclear reactors . . . » 25

A. A. Kolomenskij and A. N. Lebedev – The theory of electron motion in cyclic accelerators in presence of radiation . . . . . » 43

J. Łopuszański – On the modification of Feynman's « integral-over-all-paths » method . . . . . » 61

C. O. Beasley, Jr. and W. Holladay – Elementary particle reactions . » 77

D. Ito and H. Tanaka – Theory of meson multiple production . . . . . » 91

V. S. Barašenkov, B. M. Barbašev and E. G. Bubelev – Statistical theory of particle multiple production in high energy nucleon collisions . . . » 117

E. Fiorini, C. Giori, C. Moriggia e C. Succi – Il laboratorio per raggi cosmici del « Sabbione » . . . . . » 129

M. B. Palma-Vittorelli and M. U. Palma – A microwave electron spin resonance spectrometer . . . . . » 139

Rendiconti del V Corso che, nella Villa Monastero a Varenna, dal 14 Luglio al 3 Agosto 1957, fu tenuto a cura della Scuola Internazionale di Fisica della Società Italiana di Fisica . . . . . » 157

Fine del *Supplemento* al Vol. VII, Serie X
del *Nuovo Cimento*, 1958

PROPRIETÀ LETTERARIA RISERVATA

Direttore responsabile: G. POLVANI Tipografia Compositori - Bologna

Questo Fascicolo del *Supplemento* è stato licenziato dai torchi il 31-V-1958